Appendix K
Data Management Plan for the Mediterranean Coast Network Inventory Monitoring and Monitoring Program

National Park Service
Mediterranean Coast Network

Prepared By:

Lena Lee
Mediterranean Coast Network
401 West Hillcrest Drive
Thousand Oaks, CA 91360

FINAL

30 June 2005

Effective Date: October 2005

Recommended Citation:
Lee, L. 2005. Data management plan for the Mediterranean Coast Network Inventory and Monitoring Program. USDI National Park Service. Thousand Oaks, CA. 51 pp.

Topic(s):
Information management, guidelines

Theme Keywords:
Data management, natural resource information, ecological monitoring

Placename Keywords:
California, southern California, Mediterranean Coast Network

Acronyms:

DMP	Data Management Plan
CABR	Cabrillo National Monument
CHIS	Channel Islands National Park
I&M	Inventory and Monitoring of the National Park Service
MEDN	Mediterranean Coast Network
SAMO	Santa Monica Mountains National Recreation Area

Initial Distribution:
MEDN website:

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Number Version

Appendix K Table of Contents

1 - Introduction and Background

One of the primary goals of the National Park Service (NPS) is to "improve park management through greater reliance on scientific knowledge." The ability of resource managers to make decisions about the resources within and around park boundaries depends on our knowledge of the surrounding ecosystem. These systems are constantly evolving and our knowledge of them and how they work must also evolve for us to maintain an adequate understanding of their dynamics.

Understanding the ecosystem within our parks often begins with the collection of natural resource data collected as part of stewardship projects within the National Park system. These projects include park planning, inventories of resources, short-term and long-term monitoring, restoration, control of invasive species and other species management, fire management, trail and road maintenance, law enforcement, and the communication of natural resource information to the public (interpretation). The data collected by researchers are analyzed, synthesized and used to model various aspects of the ecosystem. Results and interpretation of the analyzed data becomes information for park managers so they can make informed decisions about the vital natural resources within the park.

1.1 – Inventory and Monitoring Program Overview

The NPS Inventory and Monitoring (I&M) Program represents a long-term commitment by the NPS to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established a framework for the I&M Program, integrating natural resource monitoring and other scientific activities into the management processes of the National Park system.

The Omnibus Management Act charges the Secretary of the Interior to "continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System," and to "… assure the full and proper utilization of the results of scientific studies for park management decisions." Section 5934 of the Act requires the Secretary of the Interior to develop a program of "inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."

To carry out this mission, the NPS initiated a service-wide, natural resource I&M program encompassing 270 parks with significant natural resources. The NPS grouped these parks into 32 networks with each network tasked to:

- document existing park vertebrates and vascular plants
- develop a management based, ecological monitoring program with a written plan and protocols
- create an information management plan that encompasses all aspects of the network program

With these tasks in mind, the I&M program's long-term goals are to:

- establish natural resource inventory and monitoring standards throughout the National Park system that transcend traditional program, activity, and funding boundaries

- inventory the natural resources and park ecosystems under National Park Service stewardship
- monitor park ecosystems to provide reference points for comparisons with other, altered environments
- integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making

- share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives

A modern information management infrastructure (e.g., staffing, hardware, software) will need to be developed to achieve the last two of these goals. This infrastructure will include procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and others will be recorded, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

1.1.1 – Data Management Goals and Objectives

As the basic and most important products of scientific research, data and information represent a valuable, and often, irreplaceable resources (Michener and Brunt, 2000). Because field experiments and associated data collection are often time and budget consuming, management of data and information products plays an important role in any scientific program. Data management is an active process with the ultimate goal of ensuring and maintaining the integrity, utility, security, and availability of data. A good data management strategy will ensure that all data:

- meet acceptable accuracy standards
- are accessible to users
- are meaningful
- contain clearly defined relationships to other relevant data
- are protected from corruption, loss, and unauthorized changes
- are maintained with integrity – indefinitely

Good data management can be accomplished through data organization, documentation, and quality control procedures. For example, organization of data in well thought out directory and file structures ensures data are physically accessible to users. Good database design facilitates data exploration and analysis. Documentation of data, or metadata, including description of the extent and purpose of data; why, when, where data are collected; how and by whom data were collected; what was done to the data; and references indicating how data have been used, is critical in maintaining the usefulness and meaning of data through time. Quality control procedures are necessary to ensure data accuracy and quality beginning with data collection and continuing through data storage and archiving. Backup strategies and data access controls will protect from data loss, including corruption and unwanted changes whether inadvertent or malicious.

1.1.2 – What is Data? – Data Defined

Data consists of factual information organized for the purpose of analysis, reason, or decision-making. Scientific data contains observational information that is obtained following a specified method or protocol. It is the intention that the information, once analyzed, will contribute to the knowledge regarding the conditions, processes, and changes occurring within our ecosystem. Such information, though, must be properly collected, verified, and stored in order for it to be used and accepted by both the scientific and non-scientific community. Defining minimum standards and procedures for data managed under the auspices of the I&M program, as well as related data external to the program, will validate their use in scientific analysis and park decision-making.

The term “data” is often used in a broader sense to encompass other products that are generated alongside primary tabular and spatial data. Data and datasets (collections of similar data) can exist in a variety of conditions and states. Generally, data products fall into five categories from raw and derived data, to documents, reports, and administrative records (Table 1.1) and can come in a range of formats including *hard-copy documents* (e.g., reports, field notes, survey forms, maps, references, administrative documents), *electronic documents* (e.g., Word files, email, websites, digital images, databases, spreadsheets, tables, delimited files), *objects* (e.g., specimens, samples, photographs, slides), and *spatial data* (e.g., shapefiles, coverages, remote-sensing data).

1.2 – Mediterranean Coast Network

The Mediterranean Coast Network (MEDN) of National Parks includes Cabrillo National Monument (CABR) in San Diego, Channel Islands National Park (CHIS) off the southern California shoreline, and Santa Monica Mountains National Recreation Area (SAMO) straddling Los Angeles and Ventura Counties along the coast

Table 1.1. Categories of data products and project deliverables.

Data Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Derived (compiled) data	Relational databases, tabular data files, GIS layers, maps, species checklists
Documents	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC/NBII metadata, data design documentation, quality assurance report, catalog of specimens/ photographs
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plan, research permit/application, other critical administrative correspondence

between Santa Monica and Oxnard, California. Together, the parks comprise examples of coastal and island southern California and share a Mediterranean-type ecosystem characterized by hot dry summers, cool wet winters, and evergreen sclerophyll shrub vegetation.

The network I&M program is organized according to the general guidance received from the national I&M program and includes a permanent network coordinator and data management specialist working in consultation with the network Board of Directors (BOD), consisting of the superintendents of each park with the network coordinator and the Pacific West Region I&M Coordinator participating as ex officio members, as well as an approved charter detailing the goals and mission of the MEDN I&M program. The BOD provides oversight and ensures individual park participation in the I&M planning process. A supporting technical committee comprised of the Chiefs of Resource Science at the three parks is the primary functioning body with responsibility for general program planning and implementation. Other park resource managers may be invited to participate on the technical committee on an ad hoc basis.

As one of the first Long-term Ecological Monitoring (LTEM) prototype parks, vital signs monitoring at CHIS has been on-going since 1991. Development and implementation of an I&M framework encompassing all three parks within the MEDN began in April, 2000. More information on the MEDN I&M program, including organization and monitoring goals, can be found in the MEDN I&M Vital Signs Monitoring Plan.

To facilitate the collection and dissemination of information, the MEDN I&M program must identify, catalog, organize, structure, archive, and provide high-quality natural resource data for park researchers, managers, and administrators. The network must also secure and maintain the quality of these data for over the long-term. Because of these complex data requirements, a process must be employed for the creation, preservation, and integration of data to make them interpretable and valuable. To accomplish this task, the MEDN I&M program has developed an information management plan to address the concerns and issues related to data management.

1.2.1 – Scope of the Information Management Plan

The MEDN I&M program approach to data management is to develop a plan that is “user friendly” to a varied audience from park natural resource managers to data managers. This may include providing guidance on data management practices at a number of different levels. To facilitate this guidance, the MEDN I&M Information Management Plan was developed along three basic principles:

- keep the plan simple, flexible, and evolving
- make it useful to all – from park GIS and data management staff, to regional technical staff, resource management staff, and cooperating scientists
- include the data users in the decision

making process whenever possible

This plan presents a broad strategy to implement and maintain a system that will serve the data and information management needs of the MEDN I&M program. This plan reflects the network's commitment to the establishment, maintenance, description, accessibility, and long-term availability of high-quality data and information. It also identifies the process by which the network will develop more detailed operational guidelines and where those guidelines will be documented. By providing for the preservation of the quality and integrity of resource data collected under the program, as well as "legacy" data currently existing, this plan will ensure that data and information will be available to assist resource managers in daily operations, make informed park management decisions, and facilitate scientific exploration and research. Specifically, the MEDN I&M Information Management Plan describes how the network will:

- support I&M Program objectives
- acquire and process data
- assure data quality
- document, analyze, summarize and disseminate data and information
- maintain nationally developed data management systems
- maintain, store and archive data

This document will detail these and other data management activities, procedures, and resources needed to maintain, over the long-term, the integrity and availability of I&M and related natural resource data for the MEDN.

1.2.2 – Data Covered by the Information Management Plan

There are many potential sources of data and information that concern the condition of natural resources in the parks or network. The types of work that may generate these natural resource data include:

- inventory and monitoring studies
- protocol development pilot studies
- special focus studies done by internal staff, contractors or cooperators
- external research projects
- monitoring or research studies done by other agencies on park or adjacent lands
- resource impact evaluations related to park planning and compliance with regulations
- resource management and restoration work

Each dataset can be grouped into four major data management categories that are coordinated or managed by the MEDN I&M program:

1. **Data managed in service-wide databases**. The MEDN uses three data systems developed by the national-level I&M program (i.e., Nature Bib, NPSpecies, and Dataset Catalog – additional details on these databases can be found in Section 4).
2. **Data developed or acquired directly by the network as a result of inventory, monitoring, or other projects**. This category includes short- and long-term tabular databases as well as project-related protocols, reports, spatial data, and associated materials such as field notes and photographs provided to the MEDN by contractors or by park staff.
3. **Data that, while not developed or maintained by the MEDN, are used as data sources or provide context to other data sets**. Examples of this category include: GIS data developed by parks, other agencies or organizations; national or international taxonomic or other classification systems; and climate or hydrologic data collected by regional or national entities.
4. **Data acquired and maintained by network parks that the MEDN assists in managing**. The MEDN may provide data management assistance for high-priority data sets or those that may benefit from standardized procedures. Examples of this category include: a multi-park database for rare plant data; data sets of legacy natural resource monitoring data; and data on exotic invasive plant species.

1.2.3 – Prioritization of Data

Because I&M data will frequently be analyzed in context with other natural resource data collected outside the program, it is critical that these external datasets meets the same standards, procedures, infrastructure, and attitudes in regards to data management. However, one challenge will be prioritizing and managing workload and other resources to meet the needs of the program.

As the focus of the MEDN I&M program is on long-term monitoring and natural resource inventories, the first priority will be management of the data and information that are derived from these primary efforts. As time and resources permit, the network will work toward raising the level of data management for current projects, legacy data, and data originating outside the I&M Program. Greatest emphasis will be placed on projects that are just beginning to be developed and implemented, because inserting good data management practices into an existing project can be difficult and will generally meet with less success.

1.2.4 – Revisions

The MEDN I&M program Information Management Plan covers program needs based on information systems technology current in 2005. However, it is structured to adapt to changing technology. As changes to technology occur, changes must also be made to the plan and all associated data management documents (e.g., guidelines, SOPs). Revisions to this plan and associated documents will be made as needed to maintain a reliable and efficient data and information handling system.

2 - Data Management Roles and Responsibilities

Data management comprises more than information technology, database theories, and applications. Data management also involves the people and organizations using or managing data and information. To serve the NPS and its constituents well, MEDN staff must understand how data and information flow including their roles and responsibilities in this process. This increased emphasis on data management, analysis, and the reporting of results will require a large investment of personnel, time, and money. The MEDN I&M Program is expected to invest at least thirty percent of available resources in developing and maintaining a high-quality data management system.

This section discusses comprehensive data management roles and responsibilities that generally apply to all network activities. Each vital sign monitoring protocol and inventory study plan will contain specific instructions for assignments and tasks that nest within this overall framework. Individuals who carry out monitoring protocols and inventory study plans are responsible for reading and understanding these instructional guidelines.

2.1 – Data Stewardship

The current and expected capacity, diversity, and rate of change in information technology makes the management of any large amount of information a greater task than any individual can expect to accomplish alone. An increasing demand for more detailed, higher quality data and information about the natural resources and ecosystem functions requires a group of knowledgeable individuals working together to ensure that data are collected using appropriate methods, and that resulting datasets, reports, maps, models, and other derived products are well managed. Datasets and the presentations of these data must be credible, representative, and available for current and future needs.

For the data management program to work effectively, everyone connected in some way with the MEDN I&M program is responsible for some stewardship role in the production, analysis, management, and/or end use of the associated data. Table 2.1 summarizes these potential roles of various park and network personnel. Each of these broad categories has principal, or 'must-do', responsibilities as well as many potential secondary tasks. Everyone included must understand the responsibilities relating to the origin, quality, disposition, and use of data to ensure a successful data management program.

2.1.1 – Data Management Roles

The responsibility for the different stewardship aspects of program data is distributed among multiple parties. Table 2.2 presents a broad description of the roles and responsibilities of various park and network personnel involved in data management.

It is important to further describe the collaborative roles and responsibilities of the network data manager, project leaders, and GIS specialists. These three roles provide the basis for data management in the MEDN I&M program. In addition to these, there are several other major roles necessary to support data stewardship. They include the information technology/systems specialist, network coordinator, and data technician.

MEDN I&M Data Manager

As coordinator of data stewardship tasks, the fundamental role of the MEDN I&M data manager will be to understand and determine program and project requirements, to create and maintain data management infrastructure and standards, and to communicate and work with all responsible individuals for a specific I&M goal or objective.

The MEDN I&M data manager plays a vital role in coordinating data management efforts for natural resource projects within the network, overseeing the design, development, implementation, and

Table 2.1. Categories of data stewardship involving all network personnel.

Stewardship Category	Related Activities	Principal Jobs or Positions
Production	Creating data or information from any original or derived source. This includes recording locations, images, measurements, and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.	Project Crew Member Project Crew Leader Data/GIS Specialist or Technician
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.	Network Coordinator Ecologist or Resource Specialist
Management	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.	Network Data Manager Project Leader GIS Manager Information Technology Specialist National-evel I&M Data Manager
End Use	Obtaining and applying information contributing to the understanding and management of park resources.	Network Coordinator Park Managers and Superintendents Others

Note: Each position is listed in only one category according to overriding responsibilities. However, most positions contribute in each category.

maintenance of data infrastructure and standards within each park, as well as between parks and the national-level I&M program. The data manager works with project leaders and GIS specialists to ensure that project data meet park and program standards, including infrastructure design, long-term data integrity, security, and data availability. Specifically, the role of MEDN I&M data manager is to:

- Work with project leaders to design databases and applications that facilitate data entry and review, quality assurance and control, and analysis and exploration
- Assist the network coordinator and project leaders in data analysis and dissemination of information to various user groups (i.e. researchers, park staff, public)
- Develop and maintain the infrastructure for metadata creation and project documentation, and ensuring that the documentation of datasets (both tabular and spatial) and database design are compliant with national metadata requirements and FDGC standards
- Ensure the long-term storage and maintenance of data follows archival procedures as documented in this plan
- Contribute to the development of standard operating procedures (SOPs) for monitoring protocols
- Coordinate data entry and maintenance of national-level I&M program applications for the network as well as create and maintain project databases in accordance with best practices and current program standards
- Provide training in the theory and practice of data management, as tailored to the needs of project personnel
- Establish and implement procedures to protect sensitive data according to project needs

Table 2.2. Summary of Roles and Responsibilities

Role	Primary responsibilities related to data management
Project Crew Member	Record and verify measurements and observations based on project objectives and protocols. Document methods, procedures and anomalies.
Project Crew Leader	Supervise crew members to ensure data collection and management obligations are met, including data verification and documentation.
Data/GIS Specialist or Technician	Perform assigned level of technical data management and/or GIS activities, including data entry, data conversion, and documentation. Work on overall data quality and stewardship with project leaders, resource specialists, and the Network data manager.
Information Technology/Systems Specialist	Provide and maintain an information systems and technology foundation to support data management.
Project Leader	Oversee and direct operations, including data management requirements, for one or more Network projects. Maintain communication with project staff, Network Data Manager, and resource specialist regarding data management.
Resource Specialist	Understand the objectives of the project, the resulting data, and their scientific and management relevance. Ensure useful data are collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reportingMake decisions about data with regard to validity, utility, sensitivity, and availability. Describe, publish, release, and discuss the data and associated information products.
GIS Manager	Support park management objectives. Coordinate and integrate local GIS and resource information management with Network, Regional, and National standards and guidelines.
Network Data Manager	Provide overall Network planning, training, and operational support for the awareness, coordination, and integration of data and information management activities, including people, information needs, data, software, and hardware. Serve as Point of Contact for National Park Service database applications (NPSpecies). Coordinate internal and external data management activities.
Database Manager	Apply particular knowledge and abilities related to database software and associated application(s).
Network Coordinator	Ensure programmatic data and information management requirements are met as part of overall Network business.
I&M Data Manager (National Level)	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases.
Other End Users	These ‘information consumers’ include park managers and superintendents, researchers, staff from other agencies, and the public. End users are responsible for the appropriate use and application of data and derived products and for providing feedback for improvements.

- Collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives

It is critical that the MEDN I&M data manager work closely with project leaders to ensure overall success in data management endeavors especially when:

- Defining the scope of the project data, and creating a data structure that meets project needs
- Becoming familiar with how the data is collected, handled and used
- Reviewing quality control and quality assurance aspects of project protocols and SOP documentation
- Identifying elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick lists and conditional validation rules
- Creating a user interface that streamlines the process of data entry, review, validation, and summarization, and that is consistent with the capabilities of the project staff
- Developing automated database procedures to improve the efficiency of the data summarization and reporting process
- Making sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data.
- Ensuring regular archival of project materials
- Informing project staff of changes and advances in data management practices

Project Leaders

Project leaders are the point of contact for an I&M project, overseeing and supervising all phases of the project. They are responsible for the coordination and supervision of all phases of the project, including data collection, validation, documentation, analysis and reporting. They are also responsible for complying with the protocol methods, data management guidelines and policies, and for the final submission of all products and deliverables. Their active involvement in data management determines the quality and usefulness of the project data. Specifically, the role of the project leader is to:

- Complete project documentation describing the 'who,' 'what,' 'where,' 'when,' 'why,' and 'how' of a project
- Develop, document and implement SOPs for field data collection and data handling
- Enact and supervise quality assurance and quality control measures for the project.
- Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection.
- Supervise or perform data entry, verification and validation.
- Maintain concise explanatory documentation of all deviations from SOPs
- Ensure documentation of important details of each field data collection period
- Maintain hard copies of data forms and send original data forms to archive on a regular basis
- Work with the network coordinator to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines
- Periodically produce, store, and make available to users regular summary reports and trend analysis of data
- Act as the main point of contact concerning project and data content

In addition, the project leader will also work closely with the MEDN I&M data manager to:

- Develop quality assurance and quality control procedures specific to project operations
- Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc
- Coordinate changes to the field data forms and the user interface for the project database
- Fully document and maintain master datasets
- Identify sensitive information that requires special consideration prior to distribution
- Manage the archival process to ensure regular archival of project documentation,

original field data, databases, reports and summaries, and other products from the project
- Define how project data will be transformed from raw data into meaningful information, and to create data summary procedures to automate and standardize this process
- Identify and prioritize legacy data for conversion, convert priority data sets to a modern format
- Increase the interpretability and accessibility of existing natural resource information

GIS Specialists
GIS support at Channel Islands National Park (CHIS) and Santa Monica Mountains National Recreation Area (SAMO) consists of staff with knowledge and skills ranging from cartography, remote sensing, digitizing, GPS field collection, database development, and programming. The GIS staff manages spatial data themes associated with network I&M projects, as well as other spatial data related to the full range of park resources. They coordinate the incorporation of spatial data into the GIS and maintain and develop metadata for geographic data in accordance with applicable standards. They are responsible for sharing and disseminating GIS data throughout the park and network. The MEDN I&M program needs are articulated to the GIS staff either by the project manager or the network data manager. The GIS specialists will work in collaboration with project leaders to:

- Determine the GIS data and analysis needs for the project
- Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques
- Display, analyze and create maps from spatial data to meet project objectives
- Properly document data in compliance with spatial metadata standards

GIS specialists will also work directly with the MEDN I&M data manager to:

- Design databases and other applications for the network
- Create relationships between GIS and non-spatial data, and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data
- Establish and implement procedures to protect sensitive spatial data according to project needs
- Develop and maintain an infrastructure for metadata creation and maintenance
- Ensure that project metadata are created and comply with national and agency standards.

Information Technology/Systems
Information Technology/Systems (IT) support at CHIS and SAMO consists of staff that maintains and upgrades software and hardware systems, establishes and maintains system security, and maintains the local area network, including all digital backups. I&M program needs are articulated to the IT staff either by the project manager or the network data manager. The MEDN is also integrated into the NPS Mediterranean & Mojave IT Support Group and benefits enormously from the expertise of system administrators and computer specialists.

Network Coordinator
The MEDN I&M Network Coordinator interfaces between project managers and network data managers. Additional information on the role and function of the network coordinator can be found in the MEDN I&M Vital Signs Monitoring Plan.

Data Technician
The data technician will assist the data manager in the implementation of the data management plan, providing technical support in database development, reporting, and maintenance of data and metadata. This position will be filled as needed.

2.2 - Data Management Coordination

Collaboration among the NPS, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in National Parks. As such, the MEDN I&M program encourages coordination among participants at all levels.

A high-level of involvement in service-wide and regional databases and data management policy, and coordination with national-level I&M program data management staff and regional resource information management personnel is necessary to ensure integration of datasets and policies. It is

important that network data managers contribute to regional and national discussions regarding data standards, data integration and analysis methods, and improving data management awareness among natural resource users. Website postings, email discussion boards, and attendance at national or regional meetings will help data managers form collaborative bonds across networks to develop and realize innovative solutions, increase combined technical capabilities, and maximize the efficiency of the program.

To promote and develop workable standards and procedures that result in the integration and availability of datasets, network data managers must also work with local network personnel, park staff, and cooperators. Key contacts for the network data manager include park GIS and data managers and the project leaders for each monitoring or inventory project. Consistent and productive communication among these personnel leads to a common understanding and better synchronization of network and park data management activities. In addition, the successful development of planning materials, inventory study plans, and monitoring protocols is dependent on the involvement and input from park scientists and resource information management staff.

3 - Data Management Program Overview

By describing the progressive stages of a project and the life cycle of the resulting data, we can more easily communicate the overall objectives and specific steps of the data management process. A generalized project work flow model provides a needed framework for data management. In addition, this awareness helps us to manage the staffing resources needed to produce, maintain, and deliver quality data and information.

This section describes the general work flow characteristics of projects that produce natural resource data and gives an overview of how natural resource data are generated, processed, finalized and made available. While the work flow presented in this plan may not apply to all situations, it does address both the long- and short-term data collection efforts anticipated by the MEDN I&M program.

3.1 – Project Work Flow

From the perspective of managing workflow, there are two main types of projects:

- *Short-term projects* – which may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring,
- *Long-term projects* – which will mainly be the implemented monitoring projects central to the MEDN I&M program, but which may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects will often require a higher level of documentation, peer review and program support.

From a data management standpoint, a primary difference between short- and long-term projects is an increased need to adhere to standards for long-term projects to ensure internal compatibility over time. While the need to follow standards is still present for short-term projects, often the cost of compliance will outweigh the benefits due to the scope, budget, and level of NPS influence over the project details. Nevertheless, both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available.

Projects can be divided into five primary stages (Figure 3.1):

1. Planning and approval
2. Design and testing
3. Implementation
4. Product integration
5. Evaluation and closure

Each stage is characterized by a set of activities carried out by park and network staff involved in the project. Primary responsibility for these activities rests with different individuals according to the different phases of a project (see Section 2 for more information on the different roles and responsibilities of park and network staff).

3.1.1 – Planning and approval

Many of the preliminary decisions regarding project scope and objectives are made during the planning and approval phase. In addition, funding sources, permits and compliance are all addressed. Primary responsibility rests with project leaders and program administrators. Although this phase lacks specific data management activities, it is important that data managers remain informed of projects, especially as timelines for deliverables are finalized. All contracts, agreements and permits should include standard language that describes the formats, specifications, and timelines for project deliverables. Examples of this standard language can be found in Appendix 3.1 (in development).

3.1.2 – Design and testing

During this phase, all of the details regarding how data will be acquired, processed, analyzed,

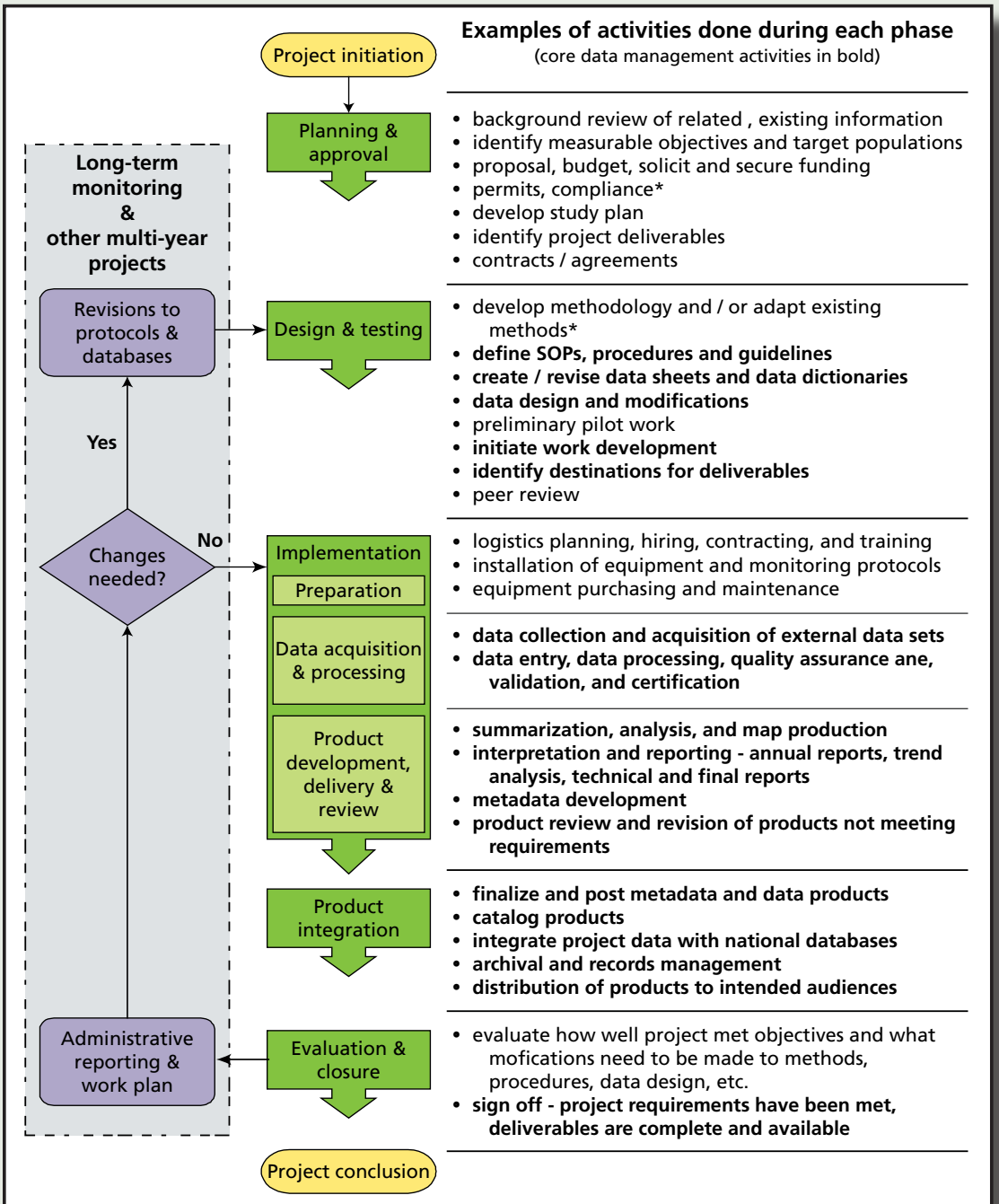

Figure 3.1. Conceptual model of project work flow (adapted from J. Boetsch, North Coast and Cascades Network)

reported and made available to others, are worked out. The project leader is responsible for developing and testing project methodology, or for modifying existing methods to meet project objectives. It is critical that the project leader and the data manager work together throughout this phase. The dialog between these two will help to build and reinforce good data management throughout the project – especially during the crucial stages of data acquisition, processing, and retrieval. By beginning collaborative development as soon as possible after project approval, data integrity and quality can most easily be assured. An important part of this collaboration is the development of the database and data dictionary, where the specifics of database implementation and parameters that will be collected are defined in detail. Devoting adequate attention to this aspect of the project is possibly the single most important part of assuring the quality, integrity and usability of the resulting data. Once the project methods, data design, and data dictionary have been developed and documented, a database can be constructed to meet project requirements.

3.1.3 – Implementation

During the implementation phase, data are acquired, processed, error-checked and documented and products, such as reports, maps, GIS themes, and other products, should be developed and delivered at this phase as well. The project leader oversees all aspects of implementation – from logistics planning, contracting, training, and equipment procurement to data acquisition, report preparation and final delivery. Throughout this phase, data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development.

Toward the end of this phase, project staff members work to develop and finalize the deliverables that were identified in the project planning documents (i.e., protocol, study plan, contract, agreement or permit). In general, all raw and derived data products, metadata, reports and other documentation must be delivered to the project leader or data manager. Administrative records must be delivered to appropriate park and network staff as specified. All project deliverables must be developed and delivered according to product specifications, which must be stipulated in all protocols, contracts, agreements, and permits.

3.1.4 – Product integration

During this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted in clearinghouses, and products are distributed or otherwise made available to their intended audience. Another aspect of integration involves the merging of certified data from a working database to a master database maintained on the network- or national-level server. Certain projects may also have additional integration needs, such as when working jointly with other agencies for a common database.

Product integration includes creating records for reports and other project documents in NatureBib, posting digital documents in the appropriate repository, posting metadata records that have been completed and submitted by project leaders, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the information from the project to be searchable and available to others via service-wide search engines.

3.1.5 – Evaluation and closure

Upon project closure, records are updated to reflect the status of the project and its associated deliverables in the MEDN I&M project tracking database (in development). For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project leaders, and data managers should work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For monitoring protocols, careful documentation of all changes is required. Changes to methods, SOPs and other procedures are maintained in a revision history log associated with each document. Major revisions may require additional peer review.

3.2 – Data Life Cycle

During various phases of a project, the data can take on different forms and be maintained in

different places as they are acquired, processed, documented and archived. This data life cycle is characterized by a sequence of events that we can model to facilitate communication. These events involve interactions with the following objects:

- *Raw data* – Analog data recorded by hand on hard-copy forms and digital files from handheld computers, GPS receivers, telemetry data loggers, etc.
- *Working database* – A project-specific database for entering and processing data for the current season (or other logical period of time). This might be the only database for short-term projects where there is no need to distinguish working data for the current season from the full set of validated project data.
- *Certified data and metadata* – Completed data and documentation for short-term projects, or one season of completed data for long-term monitoring projects. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and ready for distribution. Metadata records include the detailed information about project data needed for its proper use and interpretation (see Section 7).
- *Master database* – Project-specific database for storing the full project data set, used for viewing, summarizing, and analysis. This database is only used to store data that have passed all quality assurance steps.
- *Reports and data products* – Information that is derived from certified project data.
- *Edit log* – A means of tracking changes to certified data.
- *National databases and repositories* – Applications and repositories maintained at the national level, primarily for the purpose of integration among NPS units and for sharing information with cooperators and the public.
- *Local archives and digital library* – Local storage of copies of data, metadata and other products generated by projects. Archives are for hard-copy items and off-line storage media, whereas the digital library is maintained live on a server.

Although the data life cycle may vary depending on specific project needs and objectives, the typical life cycle for MEDN I&M projects proceeds as follows (Figure 3.2.):

1. **Acquire data** – For data recorded by hand in the field, data forms must be reviewed regularly (ideally on a daily basis) for completeness and validity in order to capture errors as close to their origin as possible.
2. **Archive raw data** – Copies of all raw data files are archived intact. Digital files are copied to the digital library section for the project; hard copy forms are either scanned and placed in the digital library, or are copied and placed in the archives. Archival or scanning of hard copy data forms may occur at the end of a season as a means of retaining all marks and edits made during the verification and validation steps.
3. **Data entry / import** – Analog data are entered manually, and digital data files are uploaded to the working database.
4. **Verification, processing and validation** – Verify accurate transcription of raw data; process data to remove missing values and other data flaws; and validate data using database queries to capture missing data, out-of-range values, and logical errors.
5. **Documentation and certification** – Develop or update project metadata and certify the dataset. Certification is a confirmation by the project leader that the data have passed all quality assurance requirements and are complete and documented. It also means that data and metadata are ready to be posted and delivered.
6. **Archive versioned dataset** – Copies of the certified data and metadata are placed in the digital library. This can be accomplished by storing a compressed copy of the working database, or by exporting data to a more software-independent format (e.g., ASCII text; see Chapter 10).
7. **Post data and update national databases** – To make data available to others, certified data and metadata are posted to national repositories such as NR-GIS Data Store. In addition, national databases such as NPSpecies, and NR-GIS Metadata Database are updated. It should be noted that data

and data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9).

8. **Upload data** – Certified data are uploaded from the working database to the master project database. This step might be skipped for short-term projects where there is no need to distinguish working data for the current season from the full set of certified project data.
9. **Reporting and analysis** – Certified data are used to generate data products, analysis, and reports, including semi-automated annual summary reports for monitoring projects. Depending on project needs, data might be exported for analysis or summarized within the database.
10. **Store products** – Reports and other data products are stored according to format and likely demand – either in the digital library, on off-line media, or in the document archives.
11. **Post products and update national databases** – To make data products available to others, reports and other products are posted to national repositories such as NR-GIS Data Store or the NR Data Image Server. In addition, products are catalogued in NatureBib. Data products may not be posted if they contain protected information about the nature or location of rare, threatened or endangered species, or other natural resources of management concern (see Chapter 9).
12. **Distribute data and information** – Data, metadata, reports and products can be shared and distributed in a variety of ways – especially via web-based national databases and repositories, by FTP or mailing in response to specific requests, or by providing direct access to project records to cooperators. In all cases, distribution will follow legal requirements under the Freedom of Information Act, (FOIA) and limitations established to protect information about sensitive resources (see Chapter 9).
13. **Track changes** – All subsequent changes to certified data are documented in a revision history, or edit, log, which accompanies project data and metadata upon distribution. Significant edits will trigger reposting of the data and products to national databases and repositories.

The sequence of events described above occurs in an iterative fashion for long-term monitoring projects, whereas for short-term projects the sequence is followed only once. For projects spanning multiple years, decision points include whether or not a separate working database is desirable, and the extent to which product development and delivery is repeated year after year.

3.3 – Sharing Data Products

Once finalized, project data and data products need to be secured in long-term storage and made available to others. To accomplish this requires a range of information systems including product repositories, clearinghouses, and web applications (Figure 3.3). The specific repositories for most MEDN I&M products are indicated in Table 3.1.

3.3.1 – Data Distribution

The process of product distribution involves several steps (Figure 3.4). As products are finalized, they can be sent to the appropriate person for integration, posting and distribution. In most cases it will be either the MEDN I&M data manager or park GIS specialist who reviews the product for consistency in format standards, then stores the product in the appropriate repository. It is expected that all products will have been reviewed for completeness and accuracy prior to delivery. After storing the products, their existence is documented by posting metadata and by updating records in the MEDN I&M project tracking application (to be developed). Metadata are then indexed by the clearinghouse function of the NR-GIS Metadata Database. These metadata records provide pointers to data and data products. Distribution then follows as data discovery allows potential users to find and either request or download the data sets from the repositories.

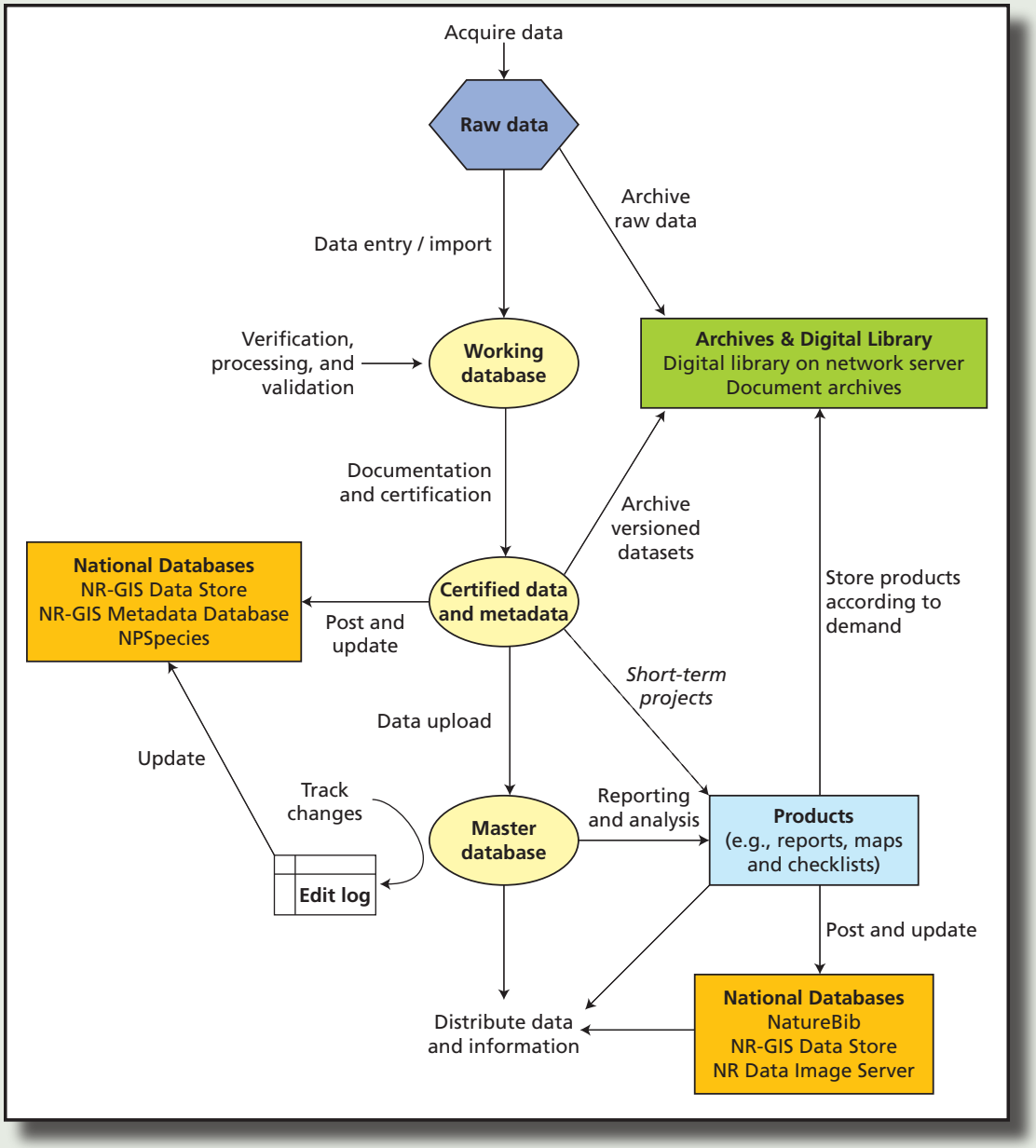

Figure 3.2. Diagram of the typical project data life cycle. (adapted from J. Boestch, North Coast and Cascades Network)

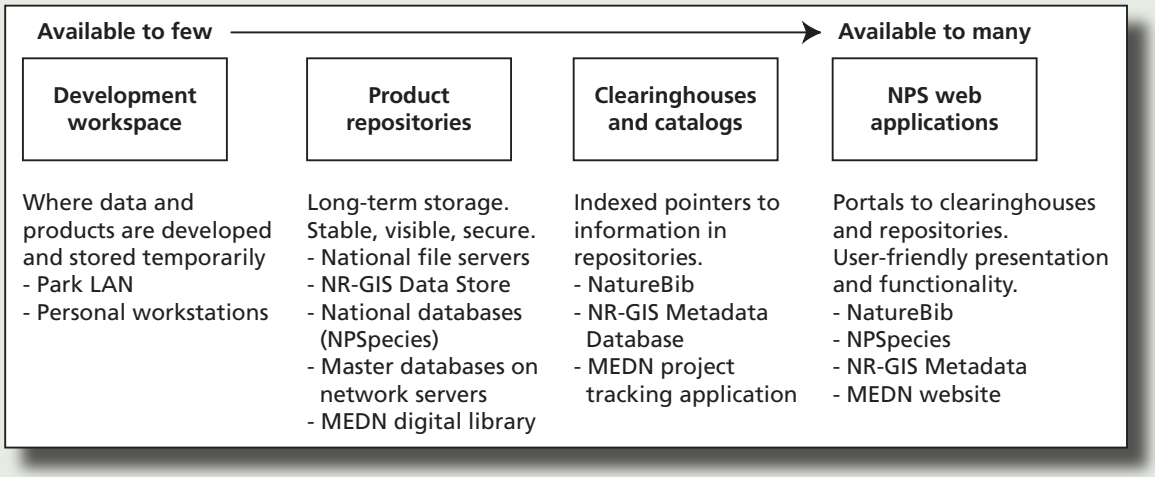

Figure 3.3. Storing and disseminating project information.

Table 3.1. Repositories for MEDN products.

Item	Repository
Reports	MEDN digital library; posted to NR Data Image Server, linked and accessed through the catalogue record in NatureBib; Park collection (hard copy)
Digital data sets (non-sensitive)	NR-GIS Data Store
Digital data, metadata, and other products Raw and finalized data Metadata, protocols, SOPs Completed reports Digital photographs, derived products	MEDN data servers, digital library, and/or other cooperators for selected monitoring projects
Project materials Vouchers specimens, raw data forms	Park archives and collections, or another specified collection
Administrative records	MEDN offices and/or park offices, park archives, National Archives

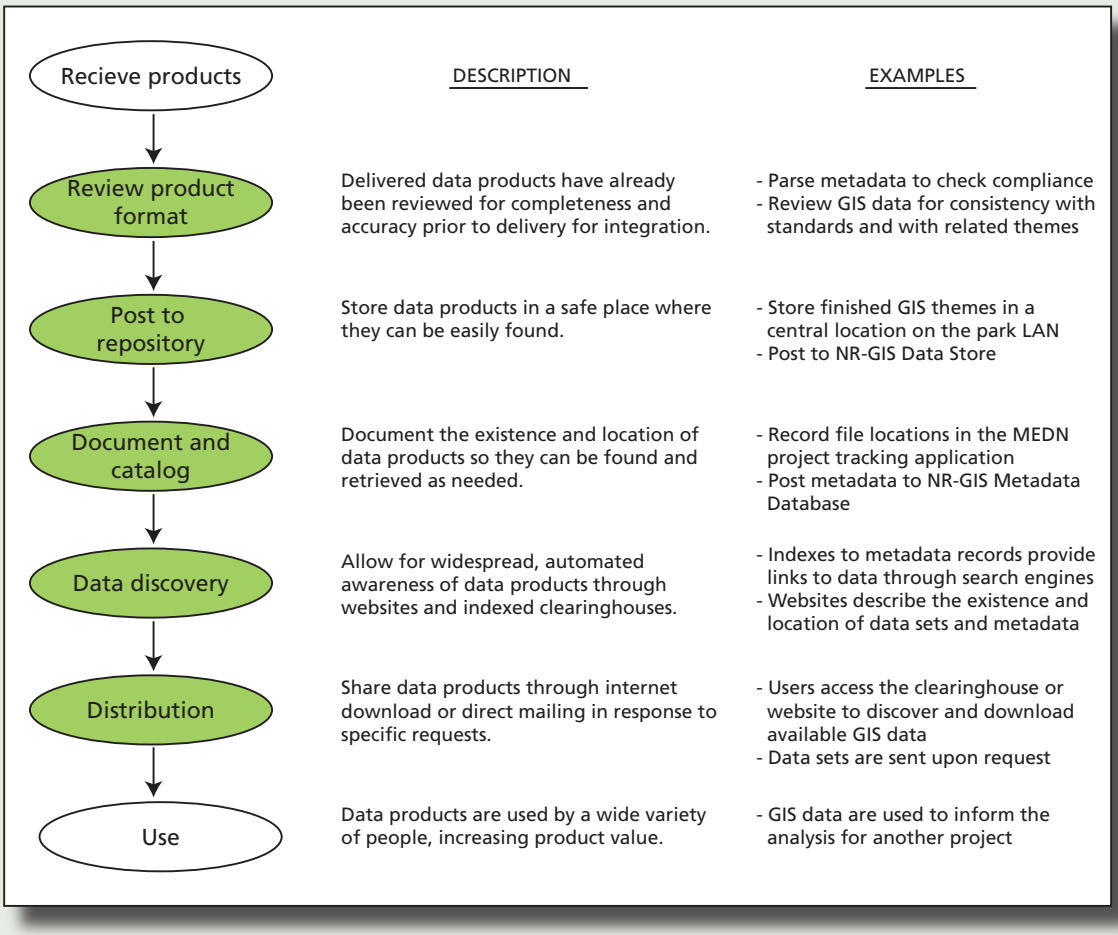

Figure 3.4. Steps involved in product distribution.

4 - Data Management Infrastructure

One of the main goals of the MEDN I&M data management program is to provide effective guidance for the management and protection of park resources as well as an efficient information systems infrastructure that integrates national-level data as well as park- and network-based data. Such a system will enhance access and use of data and information by park mangers and staff for management decisions, resource protection, and interpretation. This section will describe, in broad and general terms, the MEDN information systems infrastructure focusing specifically on the systems architecture that is central to the network data management plan.

4.1 – Data Infrastructure

As a basic foundation, a reliable, secure computer infrastructure composed of computers and servers that are functionally or directly linked through computer networking services is essential. In the MEDN, three main components make up this infrastructure. They are park-based local area networks (LAN), network data servers, and servers maintained at the national level. Each of these three components hosts different aspects of the infrastructure. For instance, national-level servers mainly host master applications (i.e., NatureBib, NPSpecies, and NR-GIS Metadata Database), centralized repositories (i.e., NR-GIS Data Store and the Protocol Clearinghouse), and public access sites (i.e, NPFocus). Network-level data servers host master project databases, including a common lookup table database (e.g. parks, projects, personnel, species, etc.) and project tracking database, as well as a network digital library for final versions of project deliverables (e.g., reports, methods documentation, data files, metadata, etc.). Park-based LANs host local and working datasets including desktop versions of national applications such as NPSpecies and Dataset Catalog, and working files of databases, geospatial themes, reports and administrative records. They are also host to the park digital library and GIS files.

4.1.1 – National-level Architecture

Information management strategies developed by the NPS Natural Resource Program Center (NRPC) and the national-level I&M program are crucial for the success of network-wide I&M projects. These strategies include an extensible desktop GIS system that integrates closely with database systems; a centralized data archiving and distribution center for long term data security and storage; sponsored training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information; and a national-level, program-wide information management framework that integrates three desktop database applications (NatureBib, NPSpecies, and Dataset Catalog) with internet-based databases to serve both local and national-level data and information requirements (Figure 4.1).

NatureBib

NatureBib is the master database for bibliographic references in the NPS. It contains citation data from independent databases such as NPSpecies, Dataset Catalog, and NR-GIS Metadata. It currently focuses on natural resource references, but will eventually be linked to references on cultural resources and other park operations. Local park or network versions of NatureBib can be downloaded from the master website into a MS Access desktop version of the application. The MS Access desktop application and additional details can be found at the NatureBib website (http://science.nature.nps.gov/im/apps/nrbib/index.htm).

NPSpecies

NPSpecies is the master database for species data in the NPS. The database contains a list of species occurring in or near each park, as well as evidence including reference to physical (e.g. voucher) or written (e.g. observations) occurrence of the species. Taxonomy and nomenclature are based on the Integrated Taxonomic Information System (ITIS). Local park or network versions

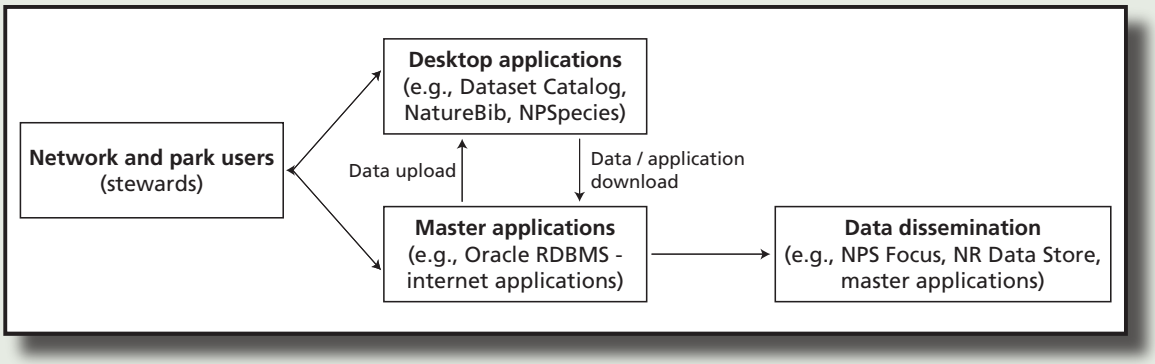

Figure 4.1. Model of the national-level application architecture.

of NPSpecies can be downloaded from the master website into the MS Access version of the application. The internet-based version of the master database is in active development and will allow park and public users to view non-sensitive data. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata to document biological inventory products. The MS Access desktop application and additional details can be found at the NPSpecies website (http://science.nature.nps.gov/im/apps/npspp/index.htm).

Dataset Catalog (and NR-GIS Metadata Database)

Dataset Catalog is a desktop metadata database application developed by the national-level I&M Program to provide a tool that parks and networks can use to inventory and manage dataset holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource datasets. Further details on its use can be found Chapter 7 of this plan. As with other service-wide applications, local park or network versions of Dataset Catalog can be downloaded from the master website (NR_GIS Metadata Database) into a MS Access desktop version of the application. References will be linked to NPSpecies and NatureBib. General information on the Dataset Catalog application and NP-GIS Metadata Database can be found at http://science.nature.nps.gov/im/apps/datacat/index.htm and http://science.nature.nps.gov/metadata, respectively.

Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database developed for MS Access for storing I&M data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS Theme Manager extension to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their I&M program and the specific sampling protocols they use. The NRDT is a key component of the national-level I&M program's standardized monitoring protocols that includes a written sampling protocol, database table structure, example data entry forms and quality checking routines, and queries and reports under development by park and network staff. Approved monitoring protocols, including the Database Template component, are made available through a web-based protocol clearinghouse (see below). A description of the Database Template application, a data dictionary, and example implementations can be found in Appendix 4.1 and on the NR Database Template website (http://science.nature.nps.gov/im/apps/template/index.htm).

Natural Resource Monitoring Protocols Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (i.e. Protocol Database) is a web-based clearinghouse of sampling protocols used in National Parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols

that have been developed by the prototype monitoring parks or other well-established protocols used within the NPS. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) for a particular protocol in MS Access that are consistent with the Natural Resource Database Template. Additional information details can be found at the Protocol Database website (http://science.nature.nps.gov/im/monitor/protocoldb.cfm).

Natural Resource Data Store

The Natural Resource (NR) Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR Data Store allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and steward their data on the server. The NR Data Store will be integrated with the master NR-GIS Metadata Database (see above) application to streamline programmatic data documentation and dissemination processes. See the NR Data Store website for further information (http://science.nature.nps.gov/nrftp).

Natural Resource – GIS Data Store

The Natural Resource (NR) – GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: http://nrdata.nps.gov/. See the NR-GIS Data Store website for further information (http://science.nature.nps.gov/nrdata).

4.1.2 – Park- and Network-level Architecture

The digital format is fast-becoming the main method of collection and storage for many of our inventory and long-term monitoring efforts. As the digital format becomes more readily available and accessible, our ability to actively maintain data will depend on our constancy in adhering to some minimum standards and configurations at the park- and network-level. In doing so, not only will continuity among datasets be achieved, but will simplify database maintenance, promote uniform working environments and facilitate the sharing of limited resources.

The MEDN I&M program will adopt a systems architecture that follows NPS and national-level I&M program structure (i.e., MS Access for databases, ArcGIS for spatial data management). Systems architecture represents the applications, database systems, repositories, and software tools that make up the framework for any data management strategy. As such, the MEDN I&M program will remain current and compatible with NPS or national-level I&M version standards for recommended applications and software programs.

The MEDN I&M program will employ a strategy that manages common tables and high-value, long-term project databases within a system that maximizes performance in a distributed, multi-user environment. Each park will host a data server that comprises the network infrastructure. The advantages of this strategy include:

- Common tables are replicated for continuous access during server down time. Also, by hosting common tables at each server, locally hosted databases can refer to the local copy of the common tables, which maximizes performance and server autonomy.
- Parallelism in data server setup simplifies system maintenance, which is important given our collaborative model for data management.
- Distribution of data services allows for program growth and balanced access loads.
- Park databases can be hosted and managed locally.

Although these three servers function as independent data nodes, each server can be accessed from any park location so long as it is within the wide area network maintained by the Pacific West Region. Final I&M program data and products will reside on one server (to be

determined) and will be synchronized to each park server as products are finalized.

4.2 – File Management Strategy

The MEDN I&M program has developed a hierarchical directory structure for storing and maintaining digital files on the network server (to be developed) and park-based LANs. Five main categories within the directory make up this structure – Administration, Databases, Libraries, Working, and GIS (Table 4.1). The key aspects of this file management strategy are to (1) separate working files from finished products, (2) make finished products largely read-only, except for "inbox" folders where users can drop things off to be cataloged and filed, and (3) enforce standards such as naming conventions and hierarchical filing within the library and database sections, though they are encouraged elsewhere as good practice. Typically, the arrangement of folders and subfolders is park-specific.

an appropriate format that is compatible with the park or network system. When well thought out, standards also help to encourage sound database design and facilitate interpretability of data sets.

All MEDN I&M projects will follow some minimum guidelines to ensure the collection of consistent data within the park and network. More detailed guidelines, such as data field names and measurement units specific to particular projects, will be included in the standard operating procedures of each project protocol. Depending on how each dataset is collected, data will be maintained either electronically or in hard copy format at each park. Much of the information to be managed may be legacy data, or data that is not current in nature. Proper storage of data (described in Chapter 10) will ensure that all data will be accessible to researchers as well as maintained over the long-term.

Generally, all tabular data generated from

Table 4.1. Recommended directory categories for storing and maintaining digital files.

Folder Name	Description of Contents
Administration	Documents related to program administration.
Databases	Local copies of national databases (e.g., NPSpecies, NatureBib, Dataset Catalog) Database(s) for common lookup tables Project back-end databases Subfolders for database development
Libraries	Read-only storage of finished project deliverables, catalogued photographs, and other reference documents generated and maintained by the program
Working	Workspace to maintain draft material and other files Note: Project leaders should work to clean out these subfolders by identifying material that belongs in one of the Libraries or elsewhere, deleting unneeded files, and moving the remainder to offline media or to a read-only folder.
GIS	Finalized base spatial data, imagery, and project-specific themes Note: This section will typically be found on a separate drive due to different storage space requirements and backup issues, and because its scope is not confined to the I&M program.

4.3 – Database Formats and Standards

Database formats and standards for all project tasks are essential for any project currently on-going or in development. They provide the guidelines necessary for data acquisition so that all data collected and produced is complete and in

inventory and monitoring studies will be stored in database format. Such databases should be designed to (1) meet the needs of the park, network and/or national program, and (2) ensure that data can be easily manipulated over time. Following common standards in database design allows the development of consistent databases,

which is extremely important when data are shared among many users, moved back and forth between different database servers and analyzed on multiple levels. Such standards include database specifications for users and developers. All data collected or generated by inventory and monitoring studies should be stored in a format that is easily exportable to commonly-used formats for analysis or manipulation.

Rather than developing a single, integrated database system, the MEDN I&M data design relies upon modular, standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project. Individual project databases and protocols can be developed at different rates to data integration. In addition, individual project databases can be modified without affecting the functionality of other project databases.
- By working with modular data sets, the network avoids a large initial investment in a centralized database and the difficulties of integrating among project areas with very different structural requirements.

4.3.1 – Centralized Database Components

Certain key information – lists of contacts, projects, parks, and species – is not only common to multiple data sets, but to the organization as a whole. It is a good strategy to centralize such information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies. Centralized information maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project management, administrative reporting, or budget management – can also link to the same tables so that all users in the network have instantaneous access to edits made by other users (Figure 4.2).

Databases developed for MEDN I&M projects will all contain the following main components:

- *Common lookup tables* – Rather than storing redundant information in each database, common tables will reside in a centralized database (Table 4.2). These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).
- *Core tables and fields based on network and national templates* – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are

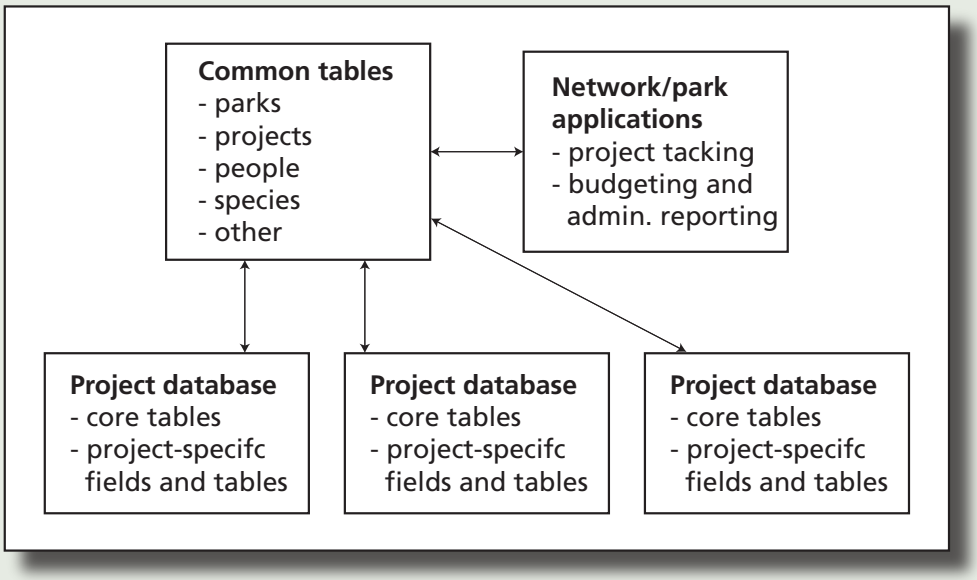

Figure 4.2. Common lookup tables and satellite databases.

populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.

- *Project-specific fields and tables* – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. For example, a temperature field will require similar data types and domain values independent of the project. Every effort will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar data sets – especially if integration with other data sets is important for meeting project objectives.

4.3.2 – Data Standards

The three types of database objects correspond to three assumed levels of data standards (Figure 4.3). Because common lookup tables are stored in one place and are referred to by multiple databases, they represent the highest level of data standard because they are implemented identically among data sets. The second level of standards is implied by the core template fields and tables, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind.

4.3.3 – National Standards Compatability

MEDN I&M program standard for fields, tables and other database specifications will mirror those conveyed through the Natural Resource Database Template as well as the NPS "Database Specifications for Inventory and Monitoring Studies" (Table 4.3.). Where there are differences between local and national standards, documentation of the rationale for these differences will be developed. In addition, documentation and database tools (e.g., queries that rename or reformat data) will be developed to ensure that data exports for integration are in a format compatible with current national standards. Standard naming conventions for databases will also follow those strongly recommended by the national-level I&M program (Table 4.4).

Table 4.2. Groupings for common lookup tables.

Grouping	Description
Parks	list of park units and networks
Projects	list of park and network projects, including inventories, monitoring, park-sponsored projects, and external research projects
People	comprehensive list of contacts for parks and network, project-specific crew lists, lists of groups and users for tracking and managing access privileges
Species	comprehensive list of taxa for the park, linkage to NPSpecies taxonomic module, project-specific species lists
Other Lookups	lists of watersheds, drainages, place names, weather conditions, habitat attributes, equipment

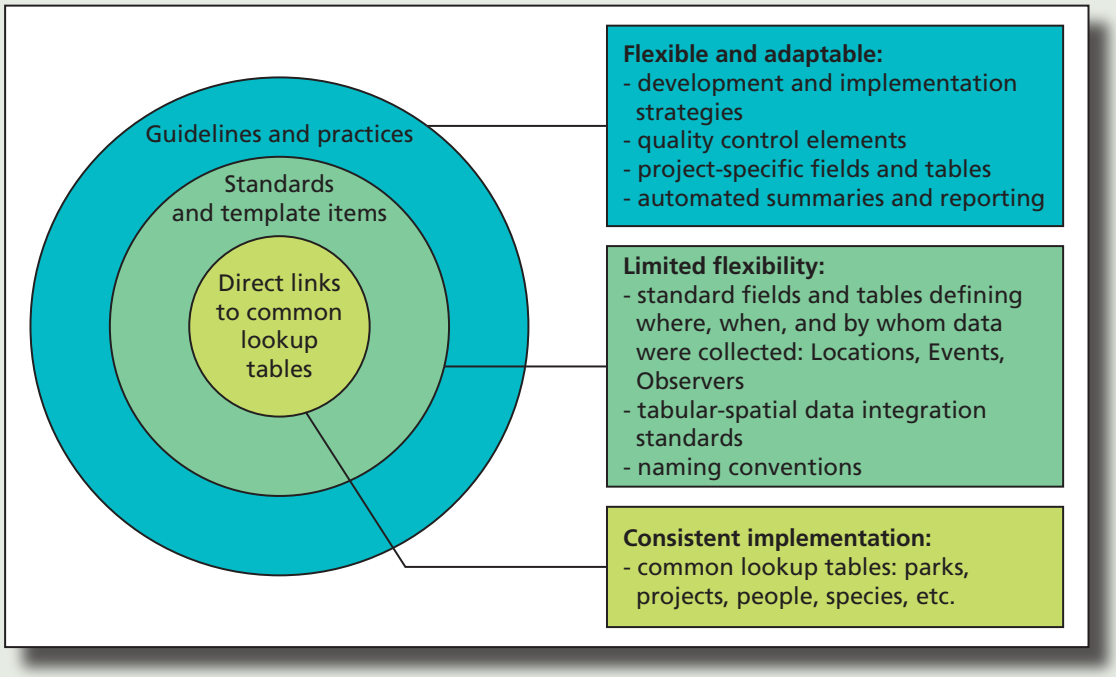

Figure 4.3. Different levels of data standards and their corresponding degree of implementation variability.

Table 4.3. Examples of database specifications for NPS I&M studies (Appendix 4.2).

Topic	Specification
Descriptive Document	Each database will be submitted with a descriptive document containing information about the project and data.
Database Template	Each database will contain the core database template tables. All other tables will be related to one of the core tables.
Data Normalization	Each database will be optimized to Third Normal Form. Each table contains data about a single subject and is identified with a primary key. No table contains repeating fields or redundant data.
Naming Standards	Each table and field name will match the corresponding standard format and adhere to the standards defined for root names.
Primary Key	Each table in the database will be identified with a primary key.
Field Description	Each field in the database will be defined by a clear and concise description.
Data Storage	Formatted text will not be stored in the database except in description or comment fields.
Required Fields	Each table in the database will be identified with required fields.
Field Domain	Each field in the database will be identified with an appropriate domain.
Field Formatting	Each field in the database may be assigned formatting options for input and/or display.
Data Verification	The data in each database will be reviewed and corrected using an approved data verification method, such that data accuracy is 95% or greater. The description of the verification method and results will be included on the Descriptive Document accompanying the database.

Table 4.4. Examples of recommended naming conventions for databases (Appendix 4.3)

Rule	Recommended Implementation
Prefix table names	Prefix each table (data, lookup, or cross-reference) with the appropriate category abbreviation. Example: data table = tbl_; lookup table = tlu_
Prefix objects	Prefix each database object with the appropriate abbreviation. Example: main form = frm_; query = qry_
Avoid spaces	Do not use spaces within a file, table or field name. Example: Instead of [MEDN SAMO]; use [MEDN_SAMO]
Separate words	Use underscores to separate words within a single table or field name.
Specific names	Choose a name that accurately identifies the data to be stored in a table or field.
Consistent case	Use mixed case text within a table or field name.
Avoid special characters	Do not use special characters in a table or field name.
Avoid unknown abbreviations	Avoid abbreviations unless necessary due to field length.
Limit length	Limit the length of table and field names to approximately 20 characters maximum.
Primary or foreign key identification	Use a standard suffix for primary and foreign keys.
Single value	Choose a field to contain a single value.
Avoid storing calculations	Choose a field to be independent of all other field values.
Singularize names	Choose the singular noun or noun-adjective form for a field name.
Avoid reserved words	Avoid a table or a field name that is a word reserved for use by the database server.

5 - Data Acquisition and Processing

Data acquisition and processing involves the collection of information from both program (I&M) and non-program (both internal and external) sources. Data collected and processed through the MEDN I&M program are considered to be program data. Data collected by entities other than the MEDN I&M Program are considered non-program data. Non-program data may be internal to the NPS (i.e. data collected by an individual park or regional park program) or external (data produced by non-NPS sources such as other government agencies or academic institutions). It is important that both the data manager and project leaders are aware of all existing data sources that could likely strengthen the scientific foundation of the program.

This section details information on policies and procedures for data collection within the scope and influence of the MEDN I&M program, including data mining procedures, field studies specifications, and anticipated approaches to changes in data collection efforts. In addition, this section will also describe prioritization, acquisition and processing of data from both NPS and external (non-NPS) sources.

5.1 – Program Data Collection

The collection of I&M data, typically by park and network staff, cooperators and/or contractors, is usually connected to either a natural resource inventory or a long-term vital signs monitoring project. Natural resource inventory projects (see 12 basic inventories of the I&M program, Appendix 5.1) are intended to identify the primary resources of each park. As such, they represent an important phase in the management of park resources. Monitoring projects related to vital signs, defined as measurable early-warning signals that indicate changes that could impair the long-term health of natural systems, allow early detection of potential problems thereby allowing park managers to take steps to restore ecological health of the resources before serious damage occurs. Working in conjunction with park resource managers and academic peers, the MEDN I&M program identified # vital signs for monitoring within the network.

5.1.1 – Data Discovery

Data discovery or data mining involves the search for existing data and information relevant to the natural resources of network parks and the MEDN I&M program mission. This process usually occurs at the onset of new projects or during the development of new protocols. With so many different sources of information, searches via the internet or through local research or academic institutions, museums, and local parks are often necessary to “discover” existing data. The types of data that are generally “discovered” include bibliographies of relevant literature through NPS and other online bibliographic databases, local libraries and park archives, geographic data from local and regional agencies, park GIS programs and geographic data clearinghouses, or biological and natural resource data contained in park, network and museum collections.

Pertinent information identified during the discovery process is likely to exist in varying conditions with much information as “completed” publications or reports. Data discovery procedures, however, should not be limited to final products as interim datasets or those in a less refined format (i.e. volunteer survey) may also yield valuable information specific to the project.

As mentioned above, most data discovery efforts are often associated with the initiation of a new project however such efforts should not be conducted solely for the purposes of protocol or project development. Data discovery is a continual process and data searches at each network park should occur on a regular basis. Encouraging data sharing within the network will not only assist in data discovery but will hopefully alleviate the need to search each park’s records regularly.

All information collected during data discovery process is maintained either electronically or in hard copy format (depending on how it was collected). Any geographic datasets collected during this process must be accompanied by FGDC compliant metadata and all datasets found (geographic or otherwise) must be entered in Dataset Catalog.

5.1.2 – Field Studies

Field studies are a main component of the I&M program and essentially involve the collection of new data based on a scientific protocol or study plan. With many different tools and methods (Appendix 5.2) available for field data collection, project leaders and the network data manager must work closely together to develop the most appropriate procedures and guidelines connected with data collection, storage and maintenance of collected field data.

To streamline the data management process, the MEDN has adopted the Natural Resources Database Template (chapter 4) as the foundation for its databases and will provide assistance in database development for any program project. Whether data collected manually or electronically, field crews should enter all data into the specified database.

5.1.3 – Changes to Procedures & Protocols

In general, changes to data collecting procedures are highly discouraged unless there are acceptable, valid reasons for altering the methodologies. All problems should be identified during the design and testing stages of the project and any changes to procedures made prior to the collection of any field data. Protocols must attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseeable problems will occur and procedures/protocols will require alteration after data collection has begun. In such case, all significant changes to the protocols must be approved by the project leader who will evaluate the proposed changes and determine if additional peer review is required before accepting them.

Alteration of data collection procedures or protocols can also occur as a result of comprehensive reviews of monitoring projects. During each review, data will be analyzed to determine whether project goals have been accomplished. If it is concluded that the protocol in its present form has not yielded sufficient data for proper analysis, changes could be recommended to meet those goals. Once again, all changes must be approved by the project leader.

5.2 – Non-Program Data Collection

Two distinct kinds of non-program data were defined in the introduction of this chapter: (1) NPS and (2) non-NPS program data. Both types of data are often very relevant to the mission of the MEDN I&M program. Data may supplement network-level projects conducted by the I&M Program or pertain to methodologies or protocols that could assist network and park personnel with the development of project SOPs.

5.2.1 – Non-Program Data Collection: Internal (NPS) Sources

A large percentage of the data collected within network parks are usually collected by park personnel involved in projects initiated at the individual park level or by other NPS regional or national programs. Network parks also engage in park-level monitoring projects which produce information that is very valuable when developing network-level monitoring protocols. It is important that park, regional and network staff work closely together to ensure that information is maintained in a manner that promotes data sharing. Parks and regional programs are encouraged to enter relevant data into national park databases such as NPSpecies and NatureBib or, at the very least, forward the information to the network data manager who will ensure that the information is properly documented and stored in manner that is accessible and available to network staff.

5.2.2 – Non-Program Data Collection: External Data Sources

External data sources often provide information important to the mission of the MEDN I&M program. Locating non-programmatic data from external data sources is a more labor intensive process requiring the review of information held at external facilities such as local libraries, universities, research institutions, and museums. A large percentage of external data products are in the form of published reports or papers, however, unpublished information such as organized volunteer surveys (e.g. Breeding Bird Surveys, Christmas Bird Counts) or information relating to ambient conditions (e.g. weather or atmospheric data) provides an indication of long-term regional

trends. Data should be obtained from the agency or individual responsible for collecting and maintaining the data. The network data manager is responsible for ensuring that the information is properly documented and appropriately stored in a manner that is accessible and available to network staff.

6 - Quality Assurance and Quality Control

The view that the ecological data and related information resulting from MEDN I&M efforts are a valuable resource worthy of preservation is justified only if those data may be used with confidence, free from error and bias. Data of poor quality can result in loss of sensitivity to subtle changes and incorrect interpretations and conclusions. The potential for problems with data quality increases with the size and complexity of the dataset (Chapal & Edwards 1994). A critical goal of the MEDN I&M data management program is to ensure that I&M projects produce data of the highest possible quality upon which to base park resource management decisions and that the long-term quality and integrity of the data are maintained.

This section will briefly discuss the various quality assurance and quality control policies and mechanisms the MEDN will use to preserve data integrity, including the identification and reduction of error at all stages in the data lifecycle (e.g, data collection, data entry, verification and validation, and dissemination).

6.1 – QA/QC Overview

Quality assurance (QA), defined as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer" (Palmer 2003), is required to ensure quality in all stages of the data development process. Quality control (QC), defined as "the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer" (Palmer 2003), procedures monitor or evaluate the resulting data products.

Both mechanisms are designed to prevent data contamination, which occurs when a process or event other than the one of interest affects the value of a variable and introduces errors into a dataset. Errors of commission include those caused by data entry and transcription errors or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced upfront with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired. Errors of omission often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four activities ranging from simple and inexpensive to more sophisticated and potentially costly: 1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata, 2) checking for unusual or unreasonable patterns in data, 3) checking for comparability of values between data sets, and 4) assessing overall data quality. Much QA/QC work is related to the first activity, which begins with data design and continues through acquisition, entry, metadata development, and archiving.

6.1.1 – NPS Mandate for Quality

Although the functional lifetime of hardware and software is decreasing rapidly, data are forever. Producers and users must know and document the quality of their data. This is especially important for sharing data and is the intent of several government directives. NPS Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service was issued in 2002 to comply with these directives to ensure and maximize the quality of information disseminated by Federal agencies. The order defines 'quality' as an encompassing term comprising three key components:

- '*Objectivity*' includes two distinct elements: 1) presentation, i.e. whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context and 2) substance, a focus on ensuring accurate, usable, and reliable information.
- '*Utility*' refers to the usefulness of the information to its intended users, from the perspectives of both the office and the public.
- '*Integrity*' refers to the security of information, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification.

Order #11B also specifies that information will be developed only from reliable data sources and that it will be accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties.

6.1.2 – Goals and Objectives

The overarching goal for data quality is to ensure that a project produces data of the right type, quality, and quantity to meet the project objectives and the user's needs. Quality criteria must be set at a level proportionate to the project-specific objectives, and these criteria must indicate the level of quality acceptable for the final data product. The EPA (2003) defines data quality objectives as qualitative and quantitative statements that:

- clarify the intended use of the data,
- define the type of data needed to support the decision,
- identify the conditions under which the data are to be collected, and
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data.

Two parameters should be considered in setting a data quality goal: 1) the percent of entries that are incorrect (frequency of errors) and 2) the magnitude of the error (criticality of errors). For example, a two-digit numeric entry off by one decimal place is a significant error. On the other hand a six-digit numeric entry with the sixth digit off by one is an insignificant error, having an accuracy of up to 99.999 percent. In another case, one incorrect digit in a six-digit species number indicates a completely different species and is quite significant. Error significance, therefore, is dependent on the type of data. The overall data quality goal should be a reasonable and attainable level of quality based on the intended use of the data and the potential consequences of making a wrong decision.

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide documented procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. These procedures and protocols should include field methodologies, training for field staff, well-organized field forms, and data entry applications with simple built-in validation.

6.2 – QA/QC Mechanisms

Although specific QA/QC procedures will depend upon the individual project or monitoring protocol, some general concepts apply. The general QA/QC procedures described below were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures.

6.2.1 – Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help reduce the incidence of invalid data in the resulting data set. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality and will reduce the overall frequency and criticality of errors at subsequent stages in the data lifecycle.

Paper field notebooks or data forms have been the primary methods for ecological data collection for many years. Although paper may have advantages in terms of longevity and ease of use, it does not work well under some environmental conditions, and processing options are limited until the data are transferred to digital format.

As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers and automated data collection instruments. Regardless of the collection method, data should ideally be transferred from one form to another only once because each transfer has the potential to introduce additional errors into the data set. One transfer should result in fewer errors, provided that appropriate QA/QC measures are incorporated into the process.

Before the data collection phase of a project begins, the data manager is responsible for providing the protocols and SOPs for data collection and storage to the project leader. All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the SOPs. The project leader, in turn, will ensure that field crews understand the procedures and closely follow them in the field. If training is necessary, the data manager will work with the project leader to provide that training. Field technicians are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

6.2.2 – Data Entry

'Data entry' is the initial set of operations where raw data from paper field forms are transcribed or typed into a computerized form linked to database tables. Spreadsheets must not be used for data entry (data can be exported to a spreadsheet for later manipulations as needed). When data were gathered or stored digitally in the field (e.g. on a data logger), data entry is the transfer of data (downloading) to a file in an office computer where they can be further manipulated. Specific procedures for electronic data transfer should be detailed in the SOPs.

Superficially, getting data from field projects into the computer seems to be a fairly simple task – the process of typing it in. Nevertheless, data entry is not a trivial concern because the value of the data depends upon their accuracy. Without proper preparation and some established guidelines, the quality and integrity of the data will be debatable. Data entry is best performed by a person who is familiar with the data and ideally takes place as soon as data collection is complete. The single goal of data entry is the transcription of the data from paper records into the computer with 100% accuracy. However, since transcription errors are virtually unavoidable during data entry, they will have to be corrected during the data verification process. Observation of certain data entry guidelines, however, will minimize verification work.

The data manager, in conjunction with the project leader, should provide training in the use of the database to all users. The project leader will ensure that users entering data understand how to enter data and that they follow the specified protocols. Users are responsible for becoming familiar with the field data forms as well as the database software and should know how to open the data entry form, create a new record or edit an existing record, and exit the database properly.

6.2.3 – Verification and Validation Procedures

Data quality is appraised by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. Performing both *verification* and *validation* of data must be stressed because it is important to remember that verified data are not always valid data. Data verification checks that the digitized data match the source data, whereas data validation checks that the data make sense. It is essential that all data are validated as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a dataset suspect. It is important to remember that only the data entry and verification stages can be done by someone who is not familiar with the kinds of errors sought during validation; validation requires in-depth knowledge about the data. As a general rule, data in each project database will be reviewed and corrected using approved verification and validation methods such that data accuracy is 95% or greater. Database fields should be included to record concerns about data integrity when applicable.

The data manager establishes SOPs for verification and validation and provides them to the project leader who will ensure that the SOPs are followed. Technicians will follow the SOPs for verification of data, documenting any changes made. After verification is complete, the project leader will validate the data. The data manager

and project leader will evaluate the results of verification and validation and determine any procedural or data form revisions that may be indicated by the results. Finally, the project leader is responsible for reviewing all data products and reports before they are released outside the network.

6.2.4 – Version Control

The MEDN I&M program manages files from a multitude of sources often comprising many iterations of a particular product in multiple formats. Some of the files are complete, some are works-in-progress, and for others, the status cannot be determined. In addition to files collected by the network, the network has also generated many files, some of which fall into the 'complete', 'works-in-progress', and 'undetermined' status categories. Determining the status of a single file can be difficult, but when confronted with a series of similarly named files, the task of determining which is the most current can become impossible.

Version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files. Such changes must not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Data progresses through various lifecycle stages, and whenever a set of changes is complete, the user must save the file with a unique name. Version control is simple insurance for maintaining data integrity, and using good version control should be routine for all data handlers.

Prior to making any major changes to a file, a copy should be stored with the appropriate version number that allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis.

The data manager determines the version control method that will be used, and other network personnel are responsible for accurately designating versions for any files upon which they have worked. Park and network staff are encouraged to design and use software tools that assist in file management. For example, databases can be created that include fields to record revision history on the file. Backup routines can be built into the databases that allow for automatic file renaming and archiving. Important program files can be catalogued in a simple index or more formally tracked and archived using professionally developed version control software. Refer to Chapter 10 of this plan for additional information on file management and storage.

6.2.5 – Data Quality Review and Communication

The MEDN I&M program requires that all datasets adhere to NPS quality and standards prior to the communication and dissemination of data and information. Director's Order #11B states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable, and timely in nature.

Periodic review of data management issues is needed to oversee and improve any data quality program. These quality checks promote a cyclic process of continuous feedback and improvement of both the data and quality planning by verifying that:

- data collection and reporting requirements are being met
- data collection and reporting procedures are being followed
- verification and validation procedures are being followed
- data file structures and maintenance is clear, accurate and according to plan
- revision control of program documents and field sheets is adequate
- calibration and maintenance procedures are being followed
- seasonal and temporary staff has been trained in data management practice
- metadata collection and construction for the program is complete
- data is being archived and catalogued appropriately for long term storage

Quality assurance procedures may need revision to improve the quality level if random checks reveal an unacceptable level of data quality. However, quality checks should not be performed with the sole objective of eliminating errors; the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the

data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality as well as controlling the lack of quality.

Frequently, it may be necessary to modify field data forms to avoid common mistakes. With knowledge of validation errors and exploratory data results in hand, the field data forms as the source of the logic errors can be reevaluated. Minor changes, small annotations, or adding check boxes to a field form remove ambiguity about what to enter on the form. In fact, any time the same type of validation error occurs repeatedly in different datasets, the field form – not the field crew – is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault. These problems can then be recognized and corrected.

The MEDN I&M program will establish guidelines and protocols to ensure compliance with DO #11B. These protocols will document quality review procedures for data and information disseminated within and outside the network, as well as a process for processing comments and complaints about data quality. Data managers must track and verify that park and network staff are operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans. Data documentation and metadata will be used to notify end users, project leaders, and network management of data quality including the results of quality assessments and the specific QA/QC procedures applied.

Any information subject for distribution must undergo internal QA/QC procedures and be approved for release. Data will be distributed to the public through the MEDN I&M website, national web sites such as the NR Data Store and the NR-GIS Data Store, and public access databases such as NPSpecies and NatureBib.

7 - Data Documentation

All datasets – whether collected last week or 20 years ago – must be accompanied by sufficient documentation (e.g., purpose and intention of the project, how and why data are collected), so that it can be reliably used well into the future. Often, data users are left with little to no information regarding the quality, completeness, or manipulations performed on a particular dataset or project. Such ambiguity results in lost productivity since the user must invest time in searching for information or render the dataset useless because answers to these and other critical questions cannot be found. Good data management will require an upfront investment in planning, organization, and documentation.

Typically, the process of documentation involves the collection of metadata. Metadata are data about the content, quality, condition and other characteristics of data. Metadata provide the means to catalog datasets within intranet and internet systems, thus making these datasets available to a broad range of potential data users. For all NPS datasets, certain directives have been issued to ensure that proper data documentation will occur. As an example, Executive Order 12906, signed by President Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) and directs agencies to plan for legacy data documentation and providing metadata and data to the public as well. Additionally, all GIS data layers must be described using the NPS Metadata Profile.

This section represents a general discussion of the process of dataset and project documentation at the network and national level. Procedures for dataset documentation include the standards and techniques to use in the production and distribution of metadata. Project documentation procedures also include methods used to track projects, including project checklists and accomplishment reports.

7.1 – Integrated Metadata System Plan and Tools

The NPS Integrated Metadata System Plan (http://science.nature.nps.gov/im/datamgmt/metaplan.htm) was designed to create an integrated, service-wide metadata database and clearinghouse system for all datasets generated, as well as a process for the dissemination of information to park staff, researchers and the public (Figure 7.1). Although there are numerous tools available for collecting metadata, three main desktop software applications are core in the creation and maintenance of metadata within this system. These are Dataset Catalog, developed by the national I&M Program, and ArcCatalog and Spatial Metadata Management System (SMMS), both commercial off-the-shelf metadata software tools. Once FGDC-compliant metadata have been generated, the metadata can be uploaded to the NPS Natural Resource – GIS Metadata and Data Store for users to access.

7.1.1 – Metadata Tools

Several software applications and components are core to the NPS Integrated Metadata System. They are Dataset Catalog, Arc Catalog and SMMS for metadata creation and maintenance, and NR – GIS Metadata and Data Store for metadata storage and dissemination.

Dataset Catalog

Developed by the national I&M program center, Dataset Catalog is a MS Access-based database tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to parks and networks. It provides parks a means to inventory, organize, and maintain information about data set holdings, and for users without

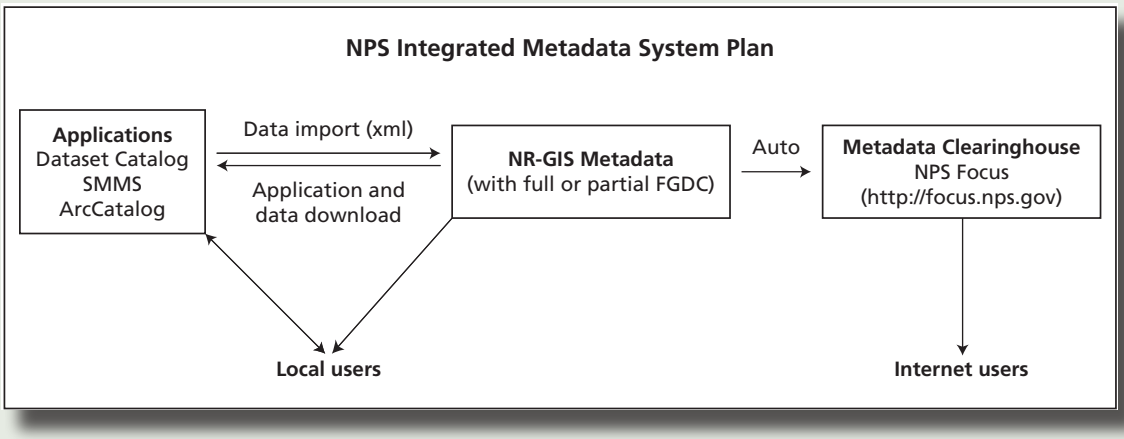

Figure 7.1. NR Integrated Metadata System (Modified from I&M Data Management Workshop, March 2004).

reliable internet connectivity, the ability to create a coherent database of their metadata holdings locally, including the capability to import and export metadata. While Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks to meet the mandates of EO 12906. Dataset Catalog has the ability to export records as an FGDC text file which can then be imported into other metadata tools. As the program is upgraded, additional capabilities including the ability to export records in Extensible Markup Languauge (XML) will added. The I&M Program recommends that all relevant datasets be cataloged, at a minimum, in Dataset Catalog format. More information on Dataset Catalog can be found at http://science.nature.nps.gov/im/apps/datacat/index.htm.

Spatial Metadata Management System (SMMS)
SMMS is a software program with the capability to create, edit, view, and publish FGDC-compliant metadata. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The NPS Integrated Metadata System Plan recommends SMMS for the FGDC Biological Data Profile. The Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Such metadata, created in compliance with the Biological Data Profile, can then be added to the National Biological Information Infrastructure (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets using SMMS is recommended by the I&M Program. More information on SMMS can be found at http://imgs.intergraph.com/smms/.

ArcCatalog
ArcCatalog is a GIS data management tool contained within the ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. With ArcCatalog users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. In addition, ArcCatalog comes with support for several popular metadata standards that allows one to create, edit, and view information about the data. Metadata records within ArcCatalog are stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata. A highly recommended extension for ArcCatalog is the NPS Metadata ArcCatalog Extension developed by NPS Midwest Region GIS Technical Support Center. The extension resolves several ArcGIS 8 metadata errors and provides added functionality for NPS users, including custom stylesheets which can be utilized to depict pertinent details more coherently than standard metadata outputs. Depending on the target audience, standardized outputs formatted using the custom stylesheets can be useful in conveying information on program data holdings and summaries of database structures. Development is also underway to provide Biological Data

Profile editing capabilities and NPS Metadata Profile support. ArcCatalog is recommended by the I&M Program for GIS-integrated geospatial metadata. More information on ArcCatalog can be found at http://www.esri.com/index.html.

Metadata Parser (mp) (optional)
The MetaParser program was developed by the US Geological Survey to compile and parse formal metadata, checking the syntax against the FGDC CSGDM and generating compliant output files for posting to clearinghouses. It also generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard. Additional information about this program can be found at http://geology.usgs.gov/tools/metadata/tools/doc/mp.html.

Natural Resource – GIS Metadata and Data Store
Until recently many NPS data stewards collected, parsed and stored metadata (and GIS data sets) in the NPS GIS Clearinghouse managed by North Carolina State University (NCSU). National-level efforts are currently underway to unify and streamline metadata development throughout the park service. This new approach, termed NR-GIS Metadata and Data Store, utilizes existing desktop metadata creation applications, an online integrated metadata database (NR-GIS Metadata), and a web-based data server (NR-GIS Data Server) to integrate both data dissemination and metadata maintenance. The system has been designed to import metadata records from the three recommended desktop applications (Dataset Catalog, SMMS, and ArcCatalog). Non-sensitive NR-GIS Metadata records are automatically posted to NPS Focus, the main gateway and search tool for NPS and public users to access NPS information resources (see Chapter 9 for more information on data distribution).

7.2 – Dataset Documentation Process

As data are obtained, whether collected in the field or mined from reference sources for each I&M project, documentation is critical to ensure validity and appropriate use of the data . The following describes the general process in the documentation and distribution of datasets in the MEDN:

1. **Identify relevant data sets and compile pertinent metadata** – The MEDN will make every effort to capture and assimilate all data relevant to the I&M program whether produced internally or externally. For each dataset obtained, metadata will be generated. In many cases, datasets will be missing pertinent information, or the originator may no longer be available, resulting in a less than adequate level of documentation. Nevertheless, every effort will be made to contact all project leaders so metadata can be acquired or a metadata interview can be conducted. A processing and revision log will be maintained with the dataset for capture of pertinent metadata.
2. **Create Dataset Catalog record** – Because metadata documentation can be accomplished in a variety of formats and levels of detail, it can become a consuming and exhaustive task. To begin documenting a dataset at the MEDN, a simple Dataset Catalog record for each relevant spatial and non-spatial dataset will be generated. This approach provides brief metadata for all network/prototype data holdings in a searchable, centralized location. In addition, managers can identify the status of metadata documentation for a particular data set (i.e., planned, in work, complete) and prioritize datasets for which formal metadata will need to be developed.
3. **Select metadata tool and complete record** – To fully document a dataset, the MEDN will select one of three metadata applications (Dataset Catalog, SMMS, or ArcCatalog) recommended by the NPS Integrated Metadata System Plan. To ensure FDGC compliance, the completed metadata will be parsed using MetaParser or similar program. Errors and omissions will be corrected until the metadata parses cleanly – indicating fully FDGC-compliance.
4. **Make information available** – Information on data holdings must be conveyed in a meaningful manner for park resource managers, researchers, and others with a potential interest in park management and research. This can be accomplished by periodically uploading

updated data and metadata to Dataset Catalog and NR_GIS Metadata. Both Dataset Catalog and NR-GIS Metadata serve as a centralized, comprehensive catalog of datasets for park staff and the public.

7.3 – Project Documentation

To support program coordination and annual reporting, and to improve accountability for the products of our natural resource inventory and monitoring efforts, the MEDN will develop and implement a project tracking database. The primary functions of this database include:

- Maintaining a single list of natural resource data-producing projects that will facilitate searches related to a specific project (e.g., status, funding sources and amounts, objectives, contact information) and summarization of the information for administrative reports.
- Maintaining a comprehensive list of what deliverables are expected and when. Once they are delivered and posted or archived, this function shifts to being a finding aid for available products. Deliverables are first specified at project initiation and information is updated at various project milestones (e.g., contracting, product delivery, archival).
- Alphanumeric codes for projects that will be used to tie together digital information in various, minimally connected systems (e.g., Research Permit and Reporting System, PMIS), along with analog materials that cannot otherwise be linked to an integrated information system. These codes can be used to link to data in databases and GIS themes, especially where information from multiple sources are stored together.

Once developed, this database will be hosted by each network park and will be replicated among parks to ensure uninterrupted access to MEDN users. Although primarily maintained by the MEDN data manager, the database will be available to project leaders, GIS staff, the network coordinator, and other network administrators. Each of these staff will be able to make certain changes to update information about project status, deliverable details, etc. Customized database views will be created to help project leaders keep on schedule, and to facilitate quick reporting on project status, accomplishments and delivered products.

This section will be updated once the application is implemented and additional details are available.

8 - Data Analysis and Reporting

For I&M data to be useful to park managers, decision-makers, or other interested parties, data must be analyzed, interpreted and reported in a form that speaks to the appropriate audience. The development of data products resulting from analyses will be guided by the vital sign monitoring protocols and will be contained in the full monitoring protocols. This chapter presents a general overview of how data collected by the MEDN will be analyzed. In addition to data analysis, various reports and other products of the monitoring effort will be briefly described.

8.1 –Trends and Analyses

The analysis and interpretation of field data and resulting management recommendation are often constrained by two factors:

- field data represent samples and statistics derived from them are estimates of real but unknown dimensions or relationships
- in evaluating comparable statistics from different populations using a statistical test, the power of the test is often not well understood and it may be necessary to know the ability of the test to detect a difference if a difference actually exists.

The guiding approach towards data analysis and interpretation of results, then, must utilize appropriate methods, including tests of statistical significance and associated power analyses. The confidence in any analysis depends upon several factors, the most robust of which is the power of the analysis. Power in a statistical sense is the probability that a given test will reject the null hypothesis when the specific alternative hypothesis is true. The power of a particular statistical test is most appropriately discussed during the sampling design phase and when calculated beforehand can be used to determine the number of samples necessary to achieve the desired power for the given suite of factors.

There are many approaches for analyzing ecological data. Relevant descriptive statistics (i.e., mean, standard deviation, and sample size) as well as indices calculating species richness, diversity and evenness for plant and animal communities can be employed for basic summary analyses. Methods that include regression analysis, t-tests, or analysis of variance can also be useful for evaluating trends. The statistical tests used will be dictated by the objectives of the monitoring protocol.

The MEDN I&M program will use third party statistical software (such as SAS, SPSS, and NCSS)for all statistical analysis,. Other external software includes special application software such as GS+ for geostatistical analysis and Distance 4.0 for estimating animal population abundance using distance sampling.

8.2 – Data Reporting

Once sufficient data become available for reliable analysis, periodic reporting will present, in detail, an interpretation of the results of monitoring efforts to date. Recommendations for action or non-action based upon the findings will be included in the report. All reporting will include an executive summary from which park and resource managers can withdraw specific findings and recommendations and from which interpreters can glean stories to pass along to the public. Results will also be considered for their value to the general scientific community and specific products (e.g. peer reviewed literature or program specific reports) will be prepared and disseminated.

8.2.1 – Types of Reports

There are several types of documents that result from analysis and reporting. They include annual reports, analysis and synthesis reports, scientific journal articles, and interpretation and outreach manuals. The types of reports that

will be generated for a particular project and the person(s) responsible for them are usually identified at project inception. These reports vary in the audience that it is written for as well as the types of reviews they will receive before dissemination.

Annual Reports
MEDN I&M projects actively collecting data will be required to generate an annual report each year, or when there are significant monitoring activities to document. These reports, once reviewed internally by the network, serve to inform park and network staff including administration and scientists working in the park. The major purposes of annual reports are to:

- Summarize annual data and document monitoring activities for the year
- Describe current condition of the resource
- Document changes in monitoring protocols
- Increase communication within the park and network

Analysis and Synthesis Reports
Analysis and synthesis reports provide critical insights into resource status and trends. The main audience for these reports is superintendents, park resource managers, network staff, and external scientists, who can use these reports to direct resource management efforts and regional resource analyses. The role of analysis and synthesis reports is to:

- Determine patterns and trends in the condition of resources being monitored
- Discover new characteristics of resources and correlations among resources being monitored
- Analyze data to determine amount of change that can be detected by this type and level of sampling
- Provide context, interpret data for the park within a multi-park, regional or national context
- Recommend changes to management of resources (feedback for adaptive management)

Scientific Journal Articles
The MEDN will seek to publish articles in scientific journals to communicate advances in knowledge to the scientific community. The scrutiny of putting a program's process is basic to science and one of the best ways to ensure rigor of a program's methods, analyses, and conclusions.

Interpretation and Outreach
Translating scientific information gained from monitoring programs for the public usually requires a concerted effort. The MEDN Research Learning Center (RLC) which promotes research in parks, as well as acting as a bridge between scientists and the public, can assist with the interpretation and presentation of monitoring information. The MEDN I&M coordinator and project leaders will work with the RLC program to form connections with college students, partners, and the interested public to provide information to the community in the best form possible.

8.2.2 – Automated Reporting and Data Summaries

Data managers are involved in multiple concurrent monitoring projects and the ability to provide the necessary support for analysis and reporting is constrained by the amount of time available for the task. For example, data collected from the field will often need to be integrated with additional, external data (i.e. weather, site information) and analysis queries will need to be developed to incorporate these different data. Thus, a schedule or timeline of anticipated reporting will need to be developed to give the project leader and data manager adequate time to meet project objectives.

To facilitate the timeliness of reports, much of the data summary analysis should be automated directly from the database. The project leader will work with the data manager to determine what kind of summaries and analyses will be useful for each report. The data manager will then assist in developing the necessary queries and user-interface that will allow the project leader to output the data into a format useful for the report. A periodic and annual report template should be developed to facilitate the extraction of data directly from the database. The tables and fields for the template should be worked out between the project leader and data manager. Templates will be tailored to meet the objective of each project, but presentation of primary information (i.e. descriptive statistics including mean, standard deviations, sample size; species richness and diversity indices) should be consistent between projects. In instances where the template will need to be modified to better meet the requirements of vital signs projects, the project leader will consult

with the data manager for any modifications to the database and summary reports.

9 - Data Dissemination

One of the most important goals of the I&M Program is to make natural resource I&M information available for NPS planning and management, as well as providing the public access to data and information of interest (Freedom of Information Act, 5 U.S.C. § 552). The Freedom of Information Act, or FOIA, is intended to establish a right for any person to access records and datasets that are owned or controlled by any federal agency, regardless of whether or not the federal government created the records. Some records and sensitive data are protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms, the internet, or via requests through a specified process (further information on the Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/).

Providing well-documented data in a timely manner is especially important to the success of the program. To accomplish this goal, procedures have been developed to ensure that quality natural resource data collected by the MEDN I&M program are made available to park managers, researchers, educators, and the general public. This chapter will detail policies on data ownership and restrictions, as well as how to request and access data collected by the MEDN I&M program.

9.1 – Data Ownership

Data and other research information collected by the MEDN I&M program will be subjected to specific ownership conditions as defined by the NPS. These conditions may include:

- Data ownership where all data and materials collected or generated using NPS personnel and funds become the property of the NPS.
- Prompt publication of any important findings from research and educational activities (authorship must accurately reflect the contributions of those involved).
- Data sharing where investigators must share collections, data, results, and supporting materials with other researchers whenever possible (in exceptional cases, where collections or data are sensitive or fragile, access may be limited).

To ensure that proper ownership, format, and development of network products is maintained, the MEDN I&M program has established guidelines for research funded by the network. All contracted work for the MEDN must be conducted as part of a signed collaborative agreement and include clear provisions for data ownership and sharing as defined by the National Park Service. Consistency and uniformity in the development and administration of grants and cooperative agreements is achieved through government-wide standards as established by the Office of Management and Budget (OMB). For instance, OMB Circular A-110 Section 36 ("Intangible Property") establishes property standards within cooperative agreements with higher institutions and non-profit organizations and describes administrative requirements pertinent to data and ownership.

9.2 – Restrictions to Sensitive Data

The NPS is directed to protect information about the nature and location of sensitive park resources and public access to these data can be restricted [Director's Order #66B, the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh)]. Through these regulations, information that could result in harm to natural resources, including

endangered or threatened species, can be classified as 'protected' or 'sensitive' and maybe withheld from public release. Information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, general information may be released, but the locations of specific nest sites would likely be withheld.

The following guidance for determining whether information originating from the MEDN should be protected is suggested in the draft Director's Order #66 (the final guidance may be contained in the Reference Manual 66):

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?
- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Based on the questions listed above, information may be limited or withheld regarding the nature and/or specific locations of resources recognized as 'sensitive' by the MEDN. 'Sensitive' resources include but are not limited to:

- endangered, threatened, rare, or commercially valuable National Park System resources (species and habitats)
- mineral or paleontological objects
- objects of cultural patrimony
- significant caves

All data and associated information from I&M activities must be assessed to determine their sensitivity. Management of resource information (e.g. reports, metadata, raw and manipulated spatial and non-spatial data, maps) that is sensitive or protected requires the:

- identification of potentially sensitive resources
- compilation of all records relating to those resources
- determination of what data must not be released to the public
- management and archival of those records to avoid their unintentional release

MEDN staff will work closely with the project leader(s), whether other network or park staff, or cooperating partners, to ensure that potentially sensitive park resources are carefully and clearly identified. Project leaders must specifically identify and 'flag' all potentially sensitive resources in all submitted products and consult with network staff before releasing any information of sensitive or protected nature in a public forum. They are also obligated to track information about these resources throughout the life of the project and will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format
- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service

9.3 - Data Requests

Every effort will be made to make all data collected by the MEDN I&M program that are not flagged as sensitive available at the designated repositories. However, for non-sensitive data not posted at these repositories, the MEDN will develop a data set request policy, to be detailed on the network website, for those who wish to acquire program data and information. This policy will include such things as:

- A statement about use and appropriate citation of data in resulting publications.
- Request that acknowledgement be given to the NPS I&M Program within all resulting reports and publications.

9.4 – Data Distribution

The MEDN I&M program will use a number of distribution methods (described in the following sections) to distribute information collected and developed by the network for park staff and the public. Both non-sensitive raw and manipulated data resulting from the MEDN I&M projects will be fully documented with FGDC compliant metadata and made available as soon as they are provided and verified by the project leader(s). Datasets for short-term studies (inventories) will be available shortly following the end of data collection or following publication of the investigator's results. Long-term (monitoring) studies will be available periodically or when analyses have been completed and reported on by the network. In addition, the MEDN I&M Program will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate distribution
- A complete record of data distribution/ dissemination is maintained

9.4.1 – Websites/Repositories

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Use of the internet will allow the distribution of data and information to reach a broad community of users. Information about the MEDN parks and I&M program is already available at several websites and online repositories.

For information on the park, the official park website (http://www.nps.gov/) provides a general description of park resources. The 'Nature and Science' pages offer additional generic information on the various natural resources that occur within the park boundaries.

Specific information on the MEDN I&M program, including program organization, inventory and monitoring goals, data and reports, may be accessed via the network website (http://www.nature.nps.gov/im/units/nw29/index.asp). Updates to the website will occur as needed. In addition, the Channel Islands National Park prototype monitoring program website (http://www.nature.nps.gov/im/units/chis/) has posted information on its long-term ecological monitoring program.

As part of the NPS I&M program, web-based applications and repositories have been developed to store a variety of park natural resource information. Public access to national and park records and datasets, such as NPSpecies and NatureBib, will be available for query on the park service's information clearinghouse, NPS Focus Digital Library and Research Station (http://focus.nps.gov). NPS Focus is a web-based 'browse and search' application that uses Z39.50 standard to interface with multiple NPS and cooperative data resource servers, including the NPS Data Clearinghouse. Documents, maps, and other data sets containing resource information from all sources, and their associated metadata, will also be assessable through NPS Focus. Available GIS data and associated metadata may be viewed and downloaded from NPFocus or the NPS Data Clearinghouse website (http://www.nps.gov/gis/data_info/clearinghouse.html).

9.5 – Information Analysis and Feedback

The MEDN I&M website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the network's I&M Program. A "comments and questions" link will be provided on the main page of the site for general questions and comments about the network's program and projects. A more specific "data error feedback" link will direct comments to the MEDN staff pertaining to errors found in website-accessible data. Annual progress reports will be presented to the Board of Directors and to the Technical Committee, and feedback will be expected during and following these presentations.

Effective long-term data maintenance is

10 - Data Maintenance, Storage and Archiving

inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This section will describe the procedures for the long-term management and maintenance of digital data, documents, and objects that result from MEDN I&M projects and activities. The overall goals of these procedures are to (1) avert the loss of information over time, (2) ensure that information and data are properly interpreted by a broad range of users, and (3) ensure that the information can be easily obtained and shared through future decades.

10.1 – Digital Data

Digital format is fast-becoming the common venue for collecting and storing data. Unfortunately, digital data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. To avert information loss due to technological obsolescence, maintaining digital files will require managing the ever-changing associated infrastructure of hardware, software, file formats, and storage media. As software and hardware evolve, data sets must be consistently migrated to new platforms or saved in formats that are independent of specific platforms or software (e.g. ascii delimited text files). One can expect major changes in hardware to occur every 1-2 years and in software every 1-5 years (Vogt-O'Connor 2000).

10.1.1 – Digital Data Maintenance

Digital data maintained over the long term can be categorized as one of two types: short-term data sets, for which for which data collections and modifications have been completed (e.g. inventory projects), and long-term data sets, for which data acquisition and entry will continue indefinitely (e.g. monitoring projects). Such active data sets require regular updates and conversion to current network and NPS database formats and software version standards. Both types of datasets, though, will generally follow the same requirements for long-term archiving.

Data maintenance

To help ensure that both inventory and monitoring data are usable in a wide range of applications or platforms, it is recommended that each data table comprising a data set is exported to ascii tab-delimited text files. These text files are in addition to the native version of the dataset, typically in database or spreadsheet format. Every text data file should be accompanied by metadata that describes the contents, relationships, and definitions of the associated data.

Because of the on-going nature of monitoring project data, raw data sets that are later manipulated or synthesized may need to be stored in perpetuity. Modifications to protocols will typically require complete data sets to be archived before modifications are implemented. Depending on the monitoring project, it may be necessary to preserve interim data sets (i.e. data "milestones") over the long term. Archived data sets or subsets destined for long-term archiving will be saved, whenever possible, in their native formats. Methods used to store short-term data sets previously mentioned will also be employed for long-term data sets. Data archiving requirements for ongoing projects will be spelled out in the data management SOPs for each monitoring project.

Quality control of converted data

All ascii files created from databases will undergo quality control (QC) to ensure that the number of records and fields correspond to the source data set. QC procedures will also ensure that conversion has not created errors or data loss. A second reviewer (preferably the project leader) will evaluate the ascii files and documentation to verify that tables, fields, and relations are fully explained and presented in a way that is useful to secondary users.

Databases that are converted from one version

of MS-Access to an upgraded version will require additional QC, especially if the databases are being actively used for data entry or analysis. Forms, queries, reports, and data entry should all be thoroughly tested to ensure a smooth transition to the newer software version. Data dictionaries that fully document field names, relationships, table keys and constraints must be created each time there is a significant change to the database. Data dictionaries and data management SOPs from previous versions should be archived along with the original database.

Version control
Completed and archived data sets that may be in older versions of MS-Access will need to be updated, with the goal of having no dataset more than two versions behind the current version used by the MEDN. There is the risk of losing a certain amount of performance in the process of conversion; for example, complex data entry forms or reports may not function properly in an upgraded version. To the extent possible, proper functionality of data entry forms and reports will be maintained, however, the priority will be to ensure basic table and relationship integrity.

Previous versions, as well as current versions of databases, will be saved and archived in their native format. Documentation of version updates and associated details will be part of the archive metadata document. Revision information and history will be included in tables within database files themselves. File names of the archived revisions must clearly indicate the revision number or date. Revision information and history must be included in tables within the database files as well.

Spatial data
Spatial data sets that are essential to the MEDN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats, and while shapefiles have retained all functionality, ArcInfo coverages may require conversion to an ArcGIS format if they are no longer supported. At this time, there is no practical way to save GIS data in a software- or platform-independent format.

Both uncorrected and corrected GPS data will be archived in their native format along with their corresponding GIS file.

10.1.2 – Data Storage and Archiving
Digital data need to be stored in a repository that ensures both security and accessibility to the data in perpetuity.

Directory structure for individual projects and electronic archiving
The organization and naming of folders and files must be intuitive to users unfamiliar with a specific project (see Chapter 4). Because each project will have its own variations and idiosyncrasies, a standardized structure is rarely realistic. All completed project archives, however, will include most of the following elements:

- administrative documents, such as agreements, contracts, correspondence, research permits
- program documents including the protocols, procedures, and worksheets
- final data submitted by contractors
- final data reformatted or manipulated by MEDN (e.g. data converted to NRDT format, data sets migrated to current software formats, statistical analysis and derived spreadsheets, etc.)
- final data in ascii tab-delimited text format
- final reports
- “read-me” text files including an explanation of directory contents, project metadata (and dataset catalog report), and information on changes or updates

In addition, ongoing monitoring projects will include several categories in addition to the above list:

- protocol documents
- SOPs
- conceptual or statistical models used for data interpretation
- data “snapshots” and/or raw data files used as source data
- more extensive read-me text files detailing protocol adjustments and resulting effects on data

Once final data and reports have been submitted, maintenance of interim or draft products is not necessary.

Backup procedures for digital data
The amount of MEDN data is increasing rapidly and the risk of data loss can come from a variety of sources including catastrophic events (e.g.

fire, flood), user error, hardware/software failure or corruption, security breaches, and vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss. Wile backup routines represent a significant investment in hardware, media, and staff time, they represent just a small percentage of the overall investment that has been made in program data. The data manger will work with park IT specialists to ensure that the backup systems used by the MEDN parks are properly maintaining data for the long-term.

Backup of data that reside on personal or local directories of staff are the responsibility of each staff member. Unless a consistent backup of data is carried out regularly, storage of data in this form is discouraged. All staff should copy all I&M related files from a personal or local directory onto their park or network servers where daily backups are performed.

Server backups at Santa Monica Mountains National Recreation Area (SAMO) are carried out by the SAMO IT specialist. Data residing on SAMO servers are backed up onto DDS 40 GB compressed tapes nightly. Incremental onsite backups occur automatically. A full backup of the entire server occurs every two weeks. Future plans include backup to an offsite server. Recent tapes are stored offsite; all others are stored onsite.

Server backups at Cabrillo National Monument are also carried out by the SAMO IT specialist. Data residing on CABR servers are backed up to an onsite 250 GB external USB hard drive. Similar to SAMO, incremental backups occur nightly. Full backups are scheduled to occur every two weeks. Future plans include backup to an offsite server.

Server backups at Channel Islands National Park (CHIS) are carried out by the CHIS IT specialist. All data is stored on a LAN server and is continuously backed up by RAID (redundant array of independent disks)-configured hard drives as well as daily tape backups. Tape backups are regularly archived onsite, offsite locally, and long distance (SAMO).

Data and network security
Access restrictions may be placed on I&M files at the discretion of the data manager and park IT specialist. For example, directories containing completed project data or interim versions of ongoing projects may be designated as read-only for all staff. In this way, all changes must be routed through the data manager, who is responsible for ensuring that documentation files associated with the data set are updated.

10.2 – Analog Data

Management guidelines for I&M analog materials, as well as digital materials, is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources, including final reports prepared by staff or contractors, program administrative documents, contracts and agreements, memoranda of agreement, and other documents related to MEDN administration, activities and projects, as well as physical items such as natural history specimens, photographs, or audio tapes, and their management are considered mission-critical records and are necessary for fulfillment of the NPS mission. NPS-19 further states: *Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.*

Section N in NPS-19 Appendix B provides guidelines on the archival of natural resource-related records including, specifically, the results of Inventory and Monitoring Programs. As stated, all natural resource records are considered "permanent" and are to be transferred to the National Archives when 30 years old. In addition, non-archival copies of natural resource-related materials are "potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

10.2.1 – Park Curatorial Staff
Park and museum curators within MEDN will remain an ongoing source of expertise, advice, and guidance on archiving and curatorial issues. Their role is vital in almost all projects undertaken by the network. Project leaders should involve park museum curators when projects are in development to ensure that all aspects of specimen curation or document archiving are considered and that associated expenses are included in project budgets.

10.2.3 – Analog Data Maintenance

Analog products come in a variety of formats. Maintaining such data over the long-term requires that products are stored on appropriate media that are designed for archival or long-term management. The most common types of analog products are documents, specimens and photographs. Each product will require its own method of long-term preservation.

Documents

Permanent paper documents housed in central files or archives should be stored on acid-free paper and folders. Paper documents may consist of project files, administrative documents and non-record copies of documents including meeting minutes, correspondence, memoranda of understanding, contracts and agreements, research permits, field data, interim and selected final reports produced by the program or under its auspices. In addition to maintaining these paper records, the MEDN will also maintain electronic versions of these documents when feasible.

Specimens

Specimens collected by the NPS for inventory or monitoring purposes must be properly processed and curated. The MEDN is fortunate to have local museums and repositories nearby to maintain specimen collections. Given the expertise in processing specimens and the facilities to house park as well as regional specimen collections, curation of specimens at the park level is a low priority.

Specimens that are collected under the auspices of the MEDN I&M program and processed by NPS staff will be curated at the network park in which they were collected or loaned to a repository approved by the park. It should be stated that all specimens on loan to repositories are considered property of the NPS. Specimen data required for cataloging (e.g. in ANCS+) will be provided to the curator/repository for each specimen.

Photographs

Slides are labeled using indelible pigment ink, or using laser-printed archival-quality slide labels. Slide labels will include: a unique ID, project name, photographer, photo date, a brief identification of contents (e.g., species name, plot ID), and geographic location (UTMs or description). All slides are stored in polypropylene slide sleeves at the MEDN. In addition, all slides will be scanned and saved as an appropriate image format. These electronic copies will then be used as the primary means for distribution or reproduction.

Photographs will be stored in individual polypropylene sleeves and within archival boxes. Each photo is labeled on the back, using archival-quality labels that are either laser-printed or hand-labeled, with the same information elements required for slides.

Every image, regardless of format, will have an entry into the MEDN I&M Photo Database (to be developed), where attributes such as electronic file name, keywords, project, photo description, photographer, date, and location are catalogued. All photo files and the associated photo database will be housed at the respective network park.

10.2.3 – Analog Data Storage and Archiving

While all I&M analog products will be stored not everything may be archived for the long-term. Such discretion will be dependent on the scope of collection (in development) for the I&M program. The MEDN will not manage documents at the park level, however, parks may choose to accession materials they feel necessary into their museums, incorporate them into their central files, or house them in their resource management library, as they deem appropriate.

Data storage – MEDN I&M central files

High-quality copies of park-related documents resulting from MEDN I&M projects, along with electronic versions, will be made accessible through the MEDN I&M central files. Maintaining a central location for all MEDN I&M documents will provide network and park staff, as well as researchers and cooperators, a ‘one-stop shop’ for retrieving information about a particular project. These files are temporarily located in the Planning, Science, and Resource Management Division at SAMO and will be moved to a more permanent location when appropriate funds are available.

Data archiving – SAMO Museum Research Building

The Museum Research Building (MRB) at SAMO has been designated as the primary storage facility for all analog products produced by the MEDN I&M program. The MRB has been designed with a temperature and humidity-control system that meets all museum standards set by the NPS. This repository will be used for original documents

and associated materials produced by the network, including original inventory reports and accompanying slides and maps, MEDN Phase 1, 2 and 3 reports, photographs, field notes, and permits, that are a high priority to maintain under archival conditions. Copies of these documents will also be maintained in the MEDN I&M central files, and all will have an electronic equivalent (e.g. pdf) for distribution or reproduction.

All MEDN I&M projects will receive a unique museum catalog number that will remain with the project for the life of the study. Information such as the scope of content, project purpose, and range of years will also be recorded to facilitate ANCS+ record creation and accession. All materials submitted as part of the project will be stored under this catalog number. The MEDN will also ensure that materials are stored on archival-quality materials (e.g. acid-free paper and folders, polypropylene or polyethylene slide pages).

Other network park museums

Other facilities include the park museums at Cabrillo NM and at CHIS. The data manager will work closely with program leaders and museum curators to ensure that all materials are properly stored and archived for the long-term.

References

Brundt, J. W. 2000. Data management principles, implementation and administration. Pp 25-47 in *Ecological data: Design, management and processing*, eds. William K. Michener and James W. Brunt. Blackwell Science, Malden, MA.

Chapal, Scott E. and Don Edwards. 1994. The development of a comprehensive data quality control system for long-term multivariate environmental databases. Chapter 10 in *Environmental information management and analysis*, eds. William K. Michener, James W. Brunt, and Susan G. Stafford. Taylor and Francis, London.

DeBacker, Michael D., Brian Witcher, and John R. Boetsch. 2002. Prairie Cluster Prototype Long-term Ecological Monitoring Program, Data Management Plan. National Park Service. Document available at: http://science.nature.nps.gov/im/monitor/networks/PrairieClusterDMP.pdf

Dye, Linda C. 1998. Channel Islands National Park, Data Management Plan. National Park Service, Technical Report 98-04. Document available at: http://www1.nrintra.nps.gov/chisdata/chis.htm

Edwards, D. 2000. Data quality assurance. Pp 70-91 in *Ecological data: Design, management and processing*, eds. William K. Michener and James W. Brunt. Blackwell Science, Malden, MA.

Palmer, Craig J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Chapter 8 in *Monitoring Ecosystems*, eds. David E. Busch and Joel C. Trexler. Island Press, Washington D. C.

Palmer, Craig J. and Eric B. Landis. 2002. Lake Mead National Recreation Area Resource Management Division, Quality System Management Plan for Environmental Data Collection Projects. National Park Service.

Tessler, Steven and Joe Gregson. 1997. Draft Data Management Protocol. National Park

Service. Document available at: http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm